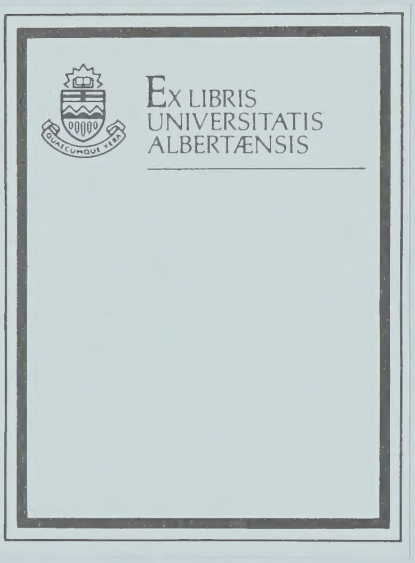

EX LIBRIS
UNIVERSITATIS
ALBERTÆNSIS

BRUCE PEEL SPECIAL COLLECTIONS LIBRARY
University of Alberta

REQUEST FOR DUPLICATION

I wish a photocopy of the thesis by Xiaohong Yao

entitled Modeling juvenile growth and mortality

The copy is for the sole purpose of private scholarly or scientific study and research. I will not reproduce, sell or distribute the copy I request, and I will not copy any substantial part of it in my own work without the permission of the copyright owner. I understand that the Library performs the service of copying at my request, and I assume all copyright responsibility for the item requested.

Date_____ Signature_____

Name_____

Address_____

List pages copied_____

Date_____ Signature_____

Digitized by the Internet Archive
in 2025 with funding from
University of Alberta Library

https://archive.org/details/0162007137894

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: **XIAOHONG YAO**

TITLE OF THESIS: **MODELING JUVENILE GROWTH AND MORTALITY RELATIONSHIPS IN MIXEDWOOD STANDS OF ALBERTA**

DEGREE: **DOCTOR OF PHILOSOPHY**

YEAR THIS DEGREE GRANTED: **1997**

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

UNIVERSITY OF ALBERTA

MODELING JUVENILE GROWTH AND MORTALITY RELATIONSHIPS IN MIXEDWOOD STANDS OF ALBERTA

BY

XIAOHONG YAO

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **DOCTOR OF PHILOSOPHY**.

DEPARTMENT OF RENEWABLE RESOURCES

EDMONTON, ALBERTA

SPRING 1997

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **MODELING JUVENILE GROWTH AND MORTALITY RELATIONSHIPS IN MIXEDWOOD STANDS OF ALBERTA** submitted by **XIAOHONG YAO** in partial fulfillment of the requirements for the degree of **DOCTOR OF PHILOSOPHY**.

To my parents

Cai, Gui Qin and Yao, Hua

Abstract

Growth and yield models play very important roles in forest management because of their abilities to update inventory, predict future yield, and explore management options and silviculture alternatives. This thesis reports on studies of large tree mortality, juvenile tree growth and mortality, and combined diameter increment of juvenile and large trees in Alberta boreal mixedwood forests for three major species: white spruce (*Picea glauca* (Moench) Voss), trembling aspen (*Populus tremuloides* Michx), and lodgepole pine (*Pinus contorta* var. *latifolia* Engelm).

Under the framework of an individual tree distance-independent growth model, a generalized logistic model of individual tree mortality was developed for mature mixedwood stands. The maximum likelihood estimator was derived in this study and implemented using the quasi-Newton algorithm. The effects of tree diameter, diameter increment, stand basal area, species composition, and site productivity on mortality were evaluated.

Basic growth and mortality relationships were developed for regenerated juvenile stands. The relationships included the tree height as a function of tree age, regeneration method, damage agents, stand density, species composition, and site productivity; the Weibull age distribution models at 2, 4, 6, 8, and 12 years since regeneration; and the logistic model relating the two-year survival probability to tree age, height, regeneration method, damage agents, stand density, species composition, and site productivity.

Using an appropriately selected unimoidal-shaped base function, and the method of parameter estimation, an annual diameter increment model was developed to predict individual tree diameter increment based on tree diameter, height, total basal area of larger trees, stand basal area, species composition, and site productivity. The selected base function prevented the model from underestimating the diameter increment for small trees, and the nonlinear fitting method

allowed data including trees with zero increment or decrement produced by natural variation in stem size.

To extend the mortality study from periodic survival to longer term survival, a lifetime distribution analysis was applied to juvenile trees. Nonparametric analysis was used to portray the survival data, and provided a reasonable theoretical model of the lifetime distribution. A parametric regression model was developed to relate the parameters of the lifetime distribution to tree age, height, regeneration method, damage agents, stand density, species composition, and site productivity.

Acknowledgments

I would like to thank all the individuals and institutions that made this study possible.

My greatest appreciation goes to my supervisor Dr. Stephen J. Titus for his continuing support, guidance, encouragement, understanding, and patience. I will be perpetually indebted to him for his acceptance of me as his student when I first came to Canada and was still in the culture shock. Throughout the entire period of my graduate study in University of Alberta, Dr. Titus has spent tremendous hours working with me in this project, and also read the manuscripts countless times and provided much needed editorial and technical review. It was fortunate for me to have this opportunity to learn from him and work with him.

I would like to gratefully acknowledge the member of my thesis supervisor committee, Dr. S. Ellen MacDonald of the Department of Renewable Resources, and Dr. Douglas P. Wiens of the Department of Mathematical Science, for reviewing the thesis draft and provided constructive criticism and guidance. Thanks are also extended to the members of my thesis examining committee, Dr. Wiktor L. Adamowicz of the Department of Rural Economy, and Dr. Bruce Dancik of the Department of Renewable Resources.

I am grateful to Dr. Lee C. Wensel of the Department of Environmental Science, Policy, and Management, University of California, Berkeley for serving as my external examiner, and his valuable, careful, and penetrating comments on the transcript. I also would like to thank Dr. Augustine C.M. Wang of Department of Mathematics, York University for his assistance in the first year of the study.

This project was supported by the Alberta Environmental Protection and Enhance Trust Fund, Alberta Land and Forest Service; the Department of Renewable Resources, University of Alberta; Pacific Forest Centre, Canadian Forest Service; and several industrial agencies. The Timber Management Branch of Alberta Land and Forest Service and Grande Prairie Operations

of Weyhaeuser Canada Ltd. provided the data for this project. Special thanks go to Mr. David J. Morgan, Mr. Tom Lakusta, and Mr. Daryl Gilday of Alberta Land and Forest Service and Mr. Greg Behuniak of Weyhaeuser Canada Ltd. for providing valuable help and cooperation. I am also grateful to Dr. Mike Bonnor of Canadian Forest Service for his interest and support.

I am grateful to the Department of Renewable Resources, University of Alberta for providing such an excellent learning environment.

I am very thankful to all my colleague friends for their assists and friendship. Particularly, Dr. Shongming Huang always instantaneously gave responses to my questions, and opened all his knowledge and experiences to me; Mr. Mike Bokalo often cheerfully helped me on computer; and Dr. Ron Hall offered me helpful advice and encouragement.

My deepest felt gratitude goes to my husband, Anwei Liu. In the past few years, no matter we were together, or geographically apart, his support has never failed. When he had to go to another city to work and this project was in the key moment, he encouraged me to stay to continue my study. When we moved together, he helped me to establish all facilities needed at home for finishing the writing of my dissertation. Even more, he often enthusiastically offered his view and suggestion on my research and sometimes even had to tolerate my impatient discussion.

Table of Contents

1. Introduction 1

2. A generalized logistic model of individual tree mortality 8

2.1 Introduction 8

2.2 Materials 9

2.2.1 Data 9

2.2.2 Response and explanatory variables 10

2.3 Methods 11

2.3.1 Model specification 11

2.3.2 Model fitting 13

2.3.3 Variable selection 18

2.3.4 Hosmer-Lemeshow goodness-of-fit test 19

2.4 Results 19

2.5 Discussion 22

2.6 References 36

3. A logistic model of juvenile tree mortality 39

3.1 Introduction 39

3.2 Materials 40

3.2.1 Data 40

3.2.2 Response and explanatory variables 41

3.3 Methods 43

3.3.1 Model specification 43

3.3.2 Maximum likelihood estimation 45

3.3.3 Variable selection 47

3.3.4 Parameter significance test 48

3.3.5 Goodness-of-fit test 48

3.3.6 Modeling strategy 49

3.4 Results 50

3.5 Discussion 52

3.5.1 White spruce model 52

3.5.2 Aspen model 53
3.5.3 Lodgepole pine model 54
3.6 References 64

4. A lifetime distribution model 66
4.1 Introduction 66
4.2 Materials 68
4.2.1 Data 68
4.2.2 Lifetime and explanatory variables 68
4.3 Methods 70
4.3.1 Nonparametric estimation 72
4.3.2 Theoretical distribution models 73
4.3.3 Parametric regression models 74
4.3.4 Maximum likelihood method 75
4.3.5 Parameter significance test 76
4.3.6 Residual analysis 77
4.4 Results and Discussion 78
4.4.1 Possible causes for mortality 78
4.4.2 Nonparametric estimation 78
4.4.3 Parametric regression model 79
4.4.4 Fitting evaluations 82
4.4.5 Model usage 83
4.5 Summary and comments 85
4.6 References 95

5. Juvenile tree height versus age relationships 97
5.1 Introduction 97
5.2 Materials and methods 98
5.2.1 Data 98
5.2.2 Explanatory variables 100
5.2.3 Model development 101
5.3 Results 102
5.3.1 Preliminary model fits 102

5.3.2 Final fits 104
5.3.3 Model testing 105
5.4 Discussion 106
5.5 References 116

6. Age distribution in regenerated stands 118
6.1 Introduction 118
6.2 Materials and methods 119
6.2.1 Data 119
6.2.2 Tree age distribution summary 120
6.2.3 Model specification 121
6.2.4 Model fitting 122
6.3 Results 124
6.4 Discussion 124
6.5 References 137

7. Individual tree diameter increment models 139
7.1 Introduction 139
7.2 Data 140
7.2.1 Permanent sample plots (PSP) 141
7.2.2 Stand dynamics (SDS) 141
7.3.3 Regenerated permanent sample plots (RPSP) 142
7.3 Model specification 143
7.4 Explanatory variables 146
7.4.1 Diameter at breast height and total height 146
7.4.2 Stand basal area 147
7.4.3 Total basal area in larger trees 147
7.4.4 Species composition 147
7.4.5 Site 148
7.5 Results 148
7.5.1 Model fitting 148
7.5.2 Residual examination 151
7.5.3 Serial correlation 151

7.5.4 Final fits 152
7.6 Discussion 152
7.7 References 163

8. General Discussion and Conclusions 166

Appendices 174
Appendix 1: Splus code for fitting the generalized logistic model based on the maximum likelihood estimation 174
Appendix 2: Splus code for Hosmer and Lemeshow goodness-of-fit test 176
Appendix 3: SAS code for fitting the logistic model of juvenile tree mortality 178
Appendix 4: SAS code for fitting the Weibull lifetime distribution model for juvenile trees 182
Appendix 5: SAS code for summarizing stand characteristics for SDS data 185
Appendix 6: SAS code for fitting the diameter increment model 189

List of Tables

Table 2-1. Summary statistics for tree and stand characteristics by life condition code 26

Table 2-2. Parameter estimates for white spruce mortality model (2.15) 27

Table 2-3. Parameter estimates for aspen mortality model (2.16) 28

Table 2-4. Parameter estimates for lodgepole pine mortality model (2.17) 29

Table 2-5. Parameter estimates for the white spruce mortality model without considering *DIN* 30

Table 2-6. Parameter estimates for the aspen mortality model without considering *DIN* 31

Table 2-7. Parameter estimates for the lodgepole pine mortality model without considering *DIN* 32

Table 3-1. Summary statistics for juvenile tree and stand characteristics by life condition code 56

Table 3-2. Parameter estimates for white spruce mortality model (3.14) 57

Table 3-3. Hosmer and Lemeshow goodness-of-fit test for white spruce model (3.14) 58

Table 3-4. Parameter estimates for aspen mortality model (3.15) 59

Table 3-5. Hosmer and Lemeshow goodness-of-fit test for aspen model (3.15) 60

Table 3-6. Parameter estimates for lodgepole pine mortality model (3.16) 61

Table 3-7. Hosmer and Lemeshow goodness-of-fit test for lodgepole pine model (3.16) 62

Table 4-1. Summary of possible mortality causes for white spruce, aspen, and lodgepole pine 87

Table 4-2. Summary statistics of the lifetime data for white spruce, aspen, and lodgepole pine 88

Table 4-3. The probability distribution function and the survival function for lifetime T 89

Table 4-4. The probability distribution function and the survival function for log lifetime Y 89

Table 4-5. Parameter estimates of the extreme value log-lifetime distribution for white spruce 90

Table 4-6. Parameter estimates of the normal log-lifetime distribution for lodgepole pine 91

Table 4-7. Parameter estimates of the normal log-lifetime distribution for aspen 92

Table 4-8. Comparison between the actual and predicted number of trees 93

Table 5-1. Summary statistics for juvenile tree and stand characteristics by species..... 109

Table 5-2. Parameter estimates for white spruce model (5.4)..... 110

Table 5-3. Parameter estimates for aspen model (5.5)..... 111

Table 5-4. Parameter estimates for lodgepole pine model (5.6)..... 112

Table 5-5. The t-test statistics for the model testing..... 113

Table 6-1. Number of plots at each year since regeneration (YSR) by species 127

Table 6-2. Trees per ha and proportional frequency at each age for white spruce 128

Table 6-3. Trees per ha and proportional frequency at each age for aspen..... 129

Table 6-4. Trees per ha and proportional frequency at each age for lodgepole pine..... 130

Table 6-5. Fit statistics for white spruce age distribution models 131

Table 6-6. Fit statistics for aspen age distribution models 132

Table 6-7. Fit statistics for lodgepole pine age distribution models..... 133

Table 7-1. Summary statistics for both mature and juvenile trees 156

Table 7-2. The parameter estimates for white spruce diameter increment model (7.10)..... 157

Table 7-3. The parameter estimates for aspen diameter increment model (7.11) 158

Table 7-4. The parameter estimates for lodgepole pine diameter increment model (7.12) 159

List of Figures

Figure 2-1. The fitted survival probability surface for aspen under different conditions of *DBH*, *DIN*, *BA*, SC_{SW}, and *SPI* 33

Figure 2-2. The fitted survival probability surface for white spruce under different conditions of *DBH*, *DIN*, *BA*, SC_{AW}, and *SPI* 34

Figure 2-3. The fitted survival probability surface for lodgepole pine under different conditions of *DBH*, *DIN*, *BA*, SC_{PL}, and *SPI* 35

Figure 3-1. The predicted survival probability in two years for white spruce seedlings, given medium site, 10000/ha aspen, and 5000/ha white spruce 63

Figure 4-1. Empirical hazard functions by species 94

Figure 4-2. Survival functions for white spruce (Example 1) 94

Figure 4-3. Survival functions for lodgepole pine (Example 2) 94

Figure 5-1. Plot of studentized residual against predicted height for white spruce height prediction model (5-4) 114

Figure 5-2. Height versus age relationships for healthy seedlings given *SITE* = medium and *DENSITY* = 15000/ha 115

Figure 6-1. Actual and predicted proportional frequency of age for white spruce 134

Figure 6-2. Actual and predicted proportional frequency of age for aspen 135

Figure 6-3. Actual and predicted proportional frequency of age for lodgepole pine 136

Figure 7-1. Plot of diameter increment against diameter for white spruce trees 160

Figure 7-2. Typical graphs of equation (7.1) by varying parameter α and β 161

Figure 7-3. Plot of studentized residual against predicted diameter increment for white spruce model (7.10) 162

Chapter 1

Introduction

Growth and yield models describing forest dynamics have been widely used in forest management because of their abilities to update inventory, predict future yield, and explore management options and silviculture alternatives. During the last two decades, along with advanced mathematical statistics and rapidly developed computing technology, growth and yield modeling methodology and technology have moved forward significantly, and many mathematically oriented growth and yield models have been constructed and employed in forest management, such as CACTOS (Wensel 1986), FOREST (Ek and Monerud 1974), MGM (Morton and Titus 1984, Huang 1992), NORM (Vanclay 1994), PROGNOSIS (Stage 1973), and PTAEDA (Daniels and Burkart 1975). The modeling unit ranges from a whole stand to an individual tree, and the modeling objective extends from stand yield prediction to biological process description. Even so, with the development of the forest industry and the changes of the forest resources, for many purposes and circumstances, we still lack adequate growth and yield information. Particularly, in Alberta boreal mixedwood forests, two questions still need to be addressed:

(1) What factors affect mortality in Alberta mixedwood stands? Mortality is difficult to predict in the mixedwood stands due to its complexity and variability. However, it is an essential feature in any growth and yield model system. A well-behaved mortality model for Alberta major species is needed.

(2) What factors influence survival and growth in regenerated juvenile stands? During the last several decades, extensive natural stands have been harvested and replaced by regenerated stands in the province. As Alberta forest industry continues to expand, our

understanding of the regenerated juvenile stand dynamics becomes increasingly important. The basic growth and mortality relationships in the juvenile stands need to be evaluated, based on which, a juvenile stand growth model can be outlined.

The boreal forests in Alberta commonly have a mixed-species composition. White spruce (*Picea glauca* (Moench) Voss), trembling aspen (*Populus tremuloides* Michx), and lodgepole pine (*Pinus contorta* var. *latifolia* Engelm) are three major species. The Mixedwood Growth Model (MGM), developed at the University of Alberta (Morton and Titus 1984, Huang 1992), is an individual tree based, distance-independent, growth simulation model. The goal of MGM is to provide logical and consistent estimates for different details in Alberta mixedwood stands. In MGM, the growth equations performed well for natural stands older than 30 years, but the mortality prediction needed to be enhanced, and the growth and yield prediction needed to be extended to the juvenile stands.

As in many individual tree growth models (Vanclay 1991, Wykoff 1990), an individual tree diameter increment model is an essential component in MGM. The model developed in Alberta mixedwood stands older than 30 years (Huang and Titus 1995) was unable to adequately predict diameter growth for juvenile trees. The available data from the juvenile stands only covered a short range of time, and the diameter increments were too variable to show a general trend. Therefore the data from the juvenile stands were merged with the data from the mature stands, and based on that, an individual tree diameter increment model was expected to work for both mature and juvenile trees.

Under the frame work of the individual tree distance-independent growth model MGM, the three primary objectives of this study are: (1) to develop an individual tree mortality model for the major species in Alberta mixedwood stands; (2) to evaluate the basic growth and mortality relationships for Alberta regenerated juvenile stands; (3) to develop an individual tree diameter increment model for both mature and juvenile trees.

To meet the first objective, the theoretical foundation for fitting the logistic model (Hosmer and Lemeshow 1989) of individual tree mortality was reviewed, and the methodology was extended to fit a generalized logistic model (Hamilton 1974, Monserud 1976) for Alberta major tree species. The generalized logistic model included a variable prediction interval and treated survival as a compound interest phenomenon. The effects of tree diameter, diameter increment, stand basal area, species composition, and site productivity on tree survival were evaluated. This analysis was based on the permanent sample plot data collected by Alberta Land and Forest Service (Alberta Forest Service 1990). Most plots were older than 30 years.

Because the interest in the regenerated stands is very recent, and it is only within about the last 15 years that permanent sample plots have begun establishment in the juvenile stands (Alberta Forest Service 1993, Gilday 1990, Weyerhaeuser Canada Ltd. 1993), the effort to meet the second objective, modeling basic relationships in the juvenile stands, had great difficulty locating appropriate data. Based on the available data, tree height versus age relationships were developed for the major species, using weighted nonlinear regression. The model expressed height as a function of tree age, regeneration method, damage agents, stand density, species composition, and site productivity. To exhibit the age structure of the newly regenerated stands, tree age distributions at 2, 4, 6, 8, 10, and 12 years since regeneration were summarized and modeled with the two-parameter Weibull distribution. Weighted nonlinear regression was used to estimate the parameters of the Weibull distribution by minimizing the weighted sum squared errors between actual and predicted proportional frequencies of trees at ages. The individual tree mortality relationships were also evaluated for the juvenile stands. Since the data contained constant remeasurement intervals, the logistic model (Hosmer and Lemeshow 1989) was used to fit the survival probability relative to tree age, height, damage agents, tree regeneration method, stand density, species composition, and site productivity.

The logistic model predicts the periodic survival probability based on tree and stand characteristics. It is easy to fit and apply. However, it has limitations in portraying the mortality process for a longer term. Even the generalized logistic model, with a variable length of the prediction interval, assumed that the lifetime of a tree is evenly distributed in the interval. The lifetime of a tree is a positive random variable, and has its own probability distribution, which may be subject to many biotic and abiotic, internal and external factors. Studying the lifetime distribution may provide a more complete picture of mortality processes. A lifetime distribution analysis (Lawless 1982) was also conducted for the juvenile stands. The intention was to provide another tool for mortality study when the periodic survival prediction is not adequate. In this study, the framework of the lifetime distribution analysis was reviewed. Nonparametric analysis was used to portray the survival data, and provided a reasonable theoretical model of the lifetime distribution. A parametric regression model was developed to predict the parameters of the lifetime distribution given tree age, height, regeneration method, damage agents, stand density, species composition, and site productivity.

To meet the third objective, an individual tree diameter increment model for both juvenile and mature trees was also developed. The data were from both juvenile stands and mature stands (Alberta Forest Service 1990, Alberta Forest Service 1993, Weyerhaeuser Canada Ltd. 1993). Using an appropriately selected unimodal-shaped diameter increment base function, and the method of parameter prediction used by Clutter et al. (1983), Huang and Titus (1995), the diameter increment model was constructed as a function of tree diameter, height, stand total basal area, total basal area of larger trees, species composition, and site productivity.

All individual models developed in this study were incorporated into the MGM model to allow growth simulation beginning with juvenile stands. The original MGM model was built based on the Permanent Sample Plots (Alberta Forest Service 1990) established in natural mixedwood stands older than 30 years. In the plots, all trees larger than 9.1 cm diameter were

remeasured, and the MGM model worked well in the range. The generalized logistic model developed in this study improved large tree mortality prediction, and the diameter increment model elaborated in this study enhanced the diameter increment prediction for both small and large trees. When tree height is less than 1.3 metre, where diameter is not defined, the height and age relationship, the age distribution model, and the logistic model of mortality developed in the juvenile stands helped to extend the MGM growth simulation to trees below 1.3 metre height.

Evaluation of predictions made by a simulation model is essential before forest managers can rely on predictions to examine implications and consequences of forest policy options and management strategies. Model validation is very important in terms of detecting faults and limitations, and avoiding unreliable predictions. For each component relationship developed here, due to the difficulty of finding an independent data set for model testing or the limited amount of data, model validation was not adequately discussed. Instead, the results of the MGM growth simulation were examined to reveal faults of the component relationships, and to find where and when the model failed. It has been a concurrent work to the model development and is still ongoing separately as a complement to the project summarized in this thesis.

This thesis is written in a paper-format with each chapter describing a separate but integrated part of the research. Three chapters (2, 3, and 4) about individual tree mortality are listed together due to their similarity in methodology and interests. Chapter 2 develops a generalized logistic model of individual tree mortality for the major tree species in Alberta mixedwood stands. Chapter 3 extends the mortality study to the juvenile stands, and develops a logistic model of juvenile tree mortality. Chapter 4 studies the lifetime distribution for juvenile trees in the regenerated stands. Chapter 5 analyzes height and age relationships in the regenerated stands. Chapter 6 summarizes the age distributions at 2, 4, 6, 8, 10, and 12 years since regeneration, and fits the distributions with the two-parameter Weibull distribution. Chapter 7

shows the development of a diameter increment model for both juvenile and mature trees. Chapter 8 provides the general discussion and conclusion.

References

Alberta Forest Service. 1990. Permanent sample plots: field procedure manual. Timber Management Branch, Alberta Forest Service, Edmonton, Alberta. FMOPC 83-03.

Alberta Forest Service 1993. Stand dynamics system field remeasurement manual. Timber Management Branch, Alberta Forest Service.

Clutter, J.L., Fortson, J.C., Pienaar, L.V., Brister, G.H., and Bailey, R.L. 1983. Timber management: a quantitative approach. John wiley & sons.

Daniels, R.F., and Burkart, H.E. 1975. Simulation of individual tree growth and stand development in managed loblolly pine plantations. Div. For. Wild. Res., Virginia Polytechnic Inst. And State University, Blacksburg, VA. FWS-5-75. 69pp.

Ek, A.R. and Monserud, R.A., 1974. Trial with program FOREST: growth and reproduction simulation for mixed species even- or uneven-aged forest stands. In J. Fries (ed.) Growth Models for Tree and Stand Simulation. Proc. IUFRO Working Party S4.01-4 meetings, 1973. Dep. For. Yield Res., Royal Coll. For., Stockholm. Res. Note 30, pp. 56-69.

Gilday, D.W. 1990. Monitor plot survey manual. Published by Alberta Forest Service.

Hamilton, D.A. 1974. Event probabilities estimated by regression. USDA Forest Service research paper. INT-152, 18p.

Hosmer, D.W. and Lemeshow, S. 1989. Applied logistic regression. John Wiley & Sons. New York. USA.

Huang, S. 1992. Diameter and height growth models. Ph.D. Thesis, Department of Forest Science, University of Alberta.

Huang, S. and Titus, S.J. 1995. An individual tree diameter increment model for white spruce in Alberta. Can. J. For. Res., 25: 1455-1465.

Lawless, J.F. 1982. Statistical models and methods for lifetime data. Wiley series in probability and mathematical statistics. John Wiley & Sons. New York.

Monserud, R.A. 1976. Simulation of forest tree mortality. Forest Science, 22:438-444.

Morton, R.T. and S.J. Titus. 1984. The development of a mixedwood stand growth model. Final report to Forest Research Branch, Alberta Forest Service.

Stage, A.R. 1973. Prognosis model for stand development. USDA For Serv. Res. Pap. INT-137.

Vanclay, J.K. 1991. Aggregating tree species to develop diameter increment equations for tropical rainforests. Forest Ecology and Management, 42: 143-168.

Vanclay. J.K. 1994. Modeling forest growth and yield - Applications to mixed tropical forests. CAB international, Wallingford, Oxon OX10 8DE, UK.

Wensel, L.C. 1986. CACTOS user's guide: the California conifer timber output simulator. Berkeley, California. Agriculture experimental station, University of California, Division of Agriculture and Natural Resources.

Weyhaeuser Canada Ltd. 1993. Permanent sample plot field procedure manual. Unpublished report. Grande Prairie Operations, Weyerhaeuser Canada Ltd., Alberta.

Wykoff, W.R. 1990. A basal area increment model for individual conifers in the northern rocky mountains. Forest Science, 36(4): 1077-1104.

Chapter 2

A generalized logistic model of individual tree mortality

2.1 Introduction

Mortality is an important component of stand development (Hamilton and Edwards 1976), and accurate prediction of tree mortality is an essential feature in any stand growth system (Monserud 1976). However, since the causes of tree death are very complex, mortality is extremely variable and difficult to predict. The earliest mortality models were stand level models. Using stand estimates for the number of trees per acre at several ages, stand level mortality functions predict the future number of trees per acre (Lee 1971, Ek 1974, Moser 1972, Somers *et al.* 1980, Clutter *et al.* 1983, Harms 1983). Individual tree mortality models are considerably different from stand level mortality models in that each individual tree involved in the growth prediction receives an estimate of the probability of survival (Clutter *et al.* 1983). Hamilton (1974) and Monserud (1976) first introduced the logistic model as an individual tree mortality function. Since then the logistic regression has been widely employed for many tree species (Hamilton and Edwards 1976, Hann 1980, Buchman and Pederson 1983, Hamilton 1986, and Hamilton 1990).

The boreal forests in Alberta commonly have a mixed-species composition, where the mortality is more difficult to predict than in pure stands. How does the mortality of major tree species in Alberta mixedwood stands relate to tree and stand characteristics? To answer this question, a model that relates the survival probability of individual trees to tree diameter at breast height, total basal area, species composition, and site productivity needs to be constructed. The available data for this study were permanent sample plots with repeated measurements for each individual tree. The remeasurement time intervals were unequal, varying from 2 to 25 years. For

the data with unequal remeasurement interval, a generalized logistic model, treating survival as a compound interest phenomenon (Hamilton 1974, Monserud 1976), was used in this study. The survival probability was modeled for three major species: trembling aspen (*Populus tremuloides* Michx), white spruce (*Picea glauca* (Moench) Voss), and lodgepole pine (*Pinus contorta* var. *latifolia* Engelm). The maximum likelihood estimation procedure was derived in this study, and accomplished by using the quasi-Newton numerical method. The likelihood ratio test was combined with our understanding of mortality process to select the important variables, and the Hosmer-Lemeshow goodness-of-fit test was conducted to evaluate the fit. Here, the performance of the model is also demonstrated, and biological interpretations are discussed.

2.2 Materials

2.2.1 Data

The Alberta Land and Forest Services provided permanent sample plot (PSP) data for this study. Since the early 1960's, over 600 permanent sample groups have been established and remeasured. Those PSP groups were randomly located throughout the inventory areas of the province to provide representative information for a variety of densities, species compositions, stand structures, and site productivities. A description of the group establishment and the data collection can be found in the field procedure manual (Alberta Forest Service 1990). Each PSP group consists of 1 or 4 plots. Most of them were beyond 30 years old at plot establishment. Since the data set was too large to be used efficiently in model development, only the first plot of each group was used. Each plot had been remeasured up to 5 times over different time intervals, varying from 2 to 25 years. At each remeasurement time, trees were measured with respect to species, diameter at breast height, and condition code. Since the data included multiple measurements for the same trees, the data were not independent, and a serial correlation among the data existed. To minimize the opportunity for the serial correlation, only two growth intervals

were used: The growth interval from the second measurement to the third measurement was selected to record the value of the explanatory variables at the beginning and observe mortality or survival response at the end, and the previous growth period, from the first measurement to the second measurement, was used to calculate the average annual diameter increment which may impact on the mortality or survival response later.

2.2.2 Response and explanatory variables

The survival response for this study was the life condition code of individuals at the end of the growth interval, which was dead or alive, a binary response. The explanatory variables should include all the measured factors that potentially influence the likelihood of mortality. Based on the variables typically recorded for PSP data, tree diameter at 1.3 m height (*DBH*), annual diameter increment (*DIN*), total basal area per ha (*BA*), species composition (SC_{SP}), and site productivity index (*SPI*) at the beginning of the growth interval, and the length of the growth interval (*L*) were considered as explanatory variables.

Tree age and height are very important factors indicating growth and the ability to resist internal and external damage agents. In this study, diameter at 1.3 m height (*DBH*) replaced age and height because tree age and height were recorded only for a small number of trees in the PSP data. Also, *DBH* is usually measured for all trees in a forest inventory project.

In order to survive, a tree must grow. The annual diameter increment (*DIN*) reflects the tree's vigor and indicates its likelihood of surviving. The *DIN* at the beginning of the growth interval was estimated from the previous remeasurement interval. It was calculated as the difference between the diameter at the first measurement and the second measurement divided by the length of the interval.

As a measure of stand competition, total basal area per ha (BA) was used, since it indicated both size and density of trees in a stand. Species composition (SC_{SP}) was defined as follows:

$$SC_{SP} = \frac{BA_{SP}}{BA},$$

where BA_{SP} was the total basal area per ha for the species under consideration, and BA was the total basal area per ha for all species.

Site index was not used in this study because age was difficult to measure accurately in the mixedwood stands and was not available for most trees. Instead, a species specific site productivity index (SPI), defined as average height of the dominant and codominant trees at 20 cm reference-diameter (Huang and Titus 1993), was used as a site productivity measure for each species in this study. The 20 cm reference-diameter corresponds roughly to the 50 year reference age in traditional Alberta site index curves (Huang and Titus 1993).

The length of growth interval was calculated as: $L = Year_2 - Year_1 + (Adj_2 - Adj_1)$, where $Year_1$ and $Year_2$ were the consecutive measurement years, Adj_1 and Adj_2 were the consecutive month adjustments. The month adjustment was defined according to the biological growth period in Alberta boreal mixedwood stands (Huang 1992). It was equal to 0.0, 0.2, 0.5, 0.9, and 1 when month was ≤ 4, $= 5$, $= 6$, $= 7$, ≥ 8, respectively. In the PSP data, the length of growth interval (L) differed from 2.2 to 24.7. For any individual tree, there is always an interval L long enough for us to observe the tree death. Therefore the survival probability model should satisfy that the survival probability is equal to 1 when L=0, and 0 when $L \rightarrow +\infty$.

2.3 Methods

2.3.1 Model specification

Assume that there are k explanatory variables $\boldsymbol{x} = (x_1, x_2, \cdots, x_k)'$ at the beginning of a growth period. L denotes the length of the growth period. The response variable y is a binary variable to indicate a tree's life condition at the end of the period, where $y = 1$ if the tree was found alive, and $y = 0$ if the tree was found dead. The plot of the response versus explanatory variables based on the data fell into two parallel lines. It did depict the dichotomous nature of the response variable clearly, but did not provide a clear picture of the relationship between survival and the explanatory variables. In this situation, a special technique, called a generalized linear model (McCullagh and Nelder 1983, Hinkley *et al.* 1991), is needed.

Let $\pi(\boldsymbol{x})$ be the mathematical expectation of y under condition $\boldsymbol{x}$, E(y|$\boldsymbol{x}$). This conditional mean is also the survival probability (p) of the individual, where $p = \text{Probability}(y = 1 | \boldsymbol{x}) = \pi(\boldsymbol{x})$. The idea of the generalized linear model is based on an assumption that a function of p may be expressed as a linear combination of $\boldsymbol{x}$. The formulation of the model is

$$g(p) = g(\pi(\boldsymbol{x})) = \beta_0 + x_1\beta_1 + x_2\beta_2 + \cdots + x_k\beta_k,$$

where $g(.)$ is a one-to-one differentiable function, called a link function, and β_0, β_1, $\cdots$, and β_k are unknown parameters (McCullagh and Nelder 1983). The link function relates the survival probability with the linear combination of the variables. Different link functions indicate different formulations of the model. The choice of a link function is usually made based on aptness to the data, model interpretation, and simplicity of the corresponding statistics (Hinkley *et al.* 1991). The most frequently used link function is the logistic link $g(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$, and the corresponding generalized linear model is called the logistic model:

$$(2.1) \qquad p = \frac{1}{1 + \exp\left[-\left(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k\right)\right]}.$$

The logistic model is also the cumulative distribution function of the logistic distribution, which is bounded by zero and one, the potential range of probabilities of survival. It has been widely used for many tree species (Hamilton and Edwards 1976, Hann 1980, Buchman and Pederson 1983, Hamilton 1986, and Hamilton 1990), since Hamilton (1974) and Monsured (1976) introduced it to model individual tree mortality. With a selection of the proper set of variables and transformations, the model should be able to describe most naturally occurring patterns of mortality (Hamilton 1990).

In this study, the growth interval L is also an explanatory variable and the model should satisfy that the p=1 when L=0, and p=0 when $L \rightarrow +\infty$. Under the assumption that, the survival time has a uniform distribution in the growth interval, Hamilton (1974) proposed a generalized logistic model which was formulated as

$$(2.2) \qquad p = \left[\frac{1}{1 + \exp\left(-\left(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k\right)\right)}\right]^L.$$

In this model, the survival probability can be projected by a variable time interval rather than a fixed yearly interval. For a live tree, at the beginning of a growth interval, i.e., L=0, the tree is definitely alive and p=1, and with increasing L, the survival probability p decreases and gradually approaches zero. We choose model (2.2) in this study. The link function for model (2.2) can be written as

$$(2.3) \qquad g\left(\pi(x)\right) = \log\left(\frac{\pi(x)^{1/L}}{1 - \pi(x)^{1/L}}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k.$$

2.3.2 Model fitting

The method to fit the generalized logistic model (2.2) is an extension of the logistic regression for model (2.1). To fit the logistic model (2.1), Hosmer and Lemeshow (1989) gave a detailed discussion, and many statistical software packages provide the specific procedures, such as the procedure PROC LOGISTIC and the procedure PROC PROBIT on SAS (SAS Institute Inc. 1992) and GLM on Splus (Statistical Science 1993). However, the estimates of parameters for model (2.2) are slightly different as explained in the following discussion.

The life condition code y is a binary response following a Bernoulli distribution when given a value of the input variable $\boldsymbol{x}$, thus for the i-th observation,

(2.4) $$\Pr ob(y_i = 1 | \boldsymbol{x}_i) = \pi(\boldsymbol{x}_i),$$

(2.5) $$\Pr ob(y_i = 0 | \boldsymbol{x}_i) = 1 - \pi(\boldsymbol{x}_i),$$

where y_i is the response variable for the i-th observation, $\boldsymbol{x}_i$ is the vector of the explanatory variable for the i-th observation. From (2.4) and (2.5), the probability distribution function of y_i can be written as:

(2.6) $$f(y_i, \pi(\boldsymbol{x}_i)) = \pi(\boldsymbol{x}_i)^{y_i} (1 - \pi(\boldsymbol{x}_i))^{1-y_i}.$$

There are n independent observations. The likelihood function is equal to

(2.7) $$L(y, \pi(\boldsymbol{x})) = \prod_{i=1}^{n} \pi(\boldsymbol{x}_i)^{y_i} (1 - \pi(\boldsymbol{x}_i))^{1-y_i}.$$

After a logarithmic transformation of (2.7), the log likelihood function is

(2.8) $$l(y, \pi(\boldsymbol{x})) = \log(L(y, \pi(\boldsymbol{x}))) = \sum_{i=1}^{n} \{ y_i \log(\pi(\boldsymbol{x}_i)) + (1 - y_i) \log(1 - \pi(\boldsymbol{x}_i)) \}.$$

From the link function (2.3), it can be derived that

(2.9) $$\pi(\boldsymbol{x}_i) = \left[1 + \exp\left(- (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}) \right) \right]^{-L_i}.$$

Substitute (2.9) into the log likelihood function (2.8). It becomes

(2.10)

$$l(y,\beta)=\sum_{i=1}^{n}\left\{-y_i L_i \log\left[1+\exp\left(-\left(\beta_0+\beta_1 x_{i1}+\cdots+\beta_k x_{ik}\right)\right)\right]+\left(1-y_i\right)\log\left[1-\left(1+\exp\left(-\left(\beta_0+\beta_1 x_{i1}+\cdots+\beta_k x_{ik}\right)\right)^{-L_i}\right)\right]\right\}$$

$$=-\sum_{\substack{i=1\\ y_i=1}}^{n} L_i \log[1+\exp(-(\beta_0+\beta_1 x_{i1}+\cdots+\beta_k x_{ik}))]+\sum_{\substack{i=1\\ y_i=0}}^{n}\log[1-(1+\exp(-(\beta_0+\beta_1 x_{i1}+\cdots+\beta_k x_{ik})))^{-L_i}].$$

The log likelihood function contains the unknown parameter β. Maximizing the log likelihood function (2.10), the maximum likelihood estimates $\hat{\beta}$ for model (2.2) can be obtained.

Different link functions provide different presentations of $\pi(x_i)$ in (2.9) and lead to different log likelihood functions. Function (2.10) is not the same as the log likelihood function for model (2.1). That is why we cannot use the procedures built for fitting model (2.1) to fit model (2.2).

In this study, the Splus statistical language (Statistical Science 1993) was used to obtain the maximum likelihood estimates for model (2.2) (Appendix 1). In Splus, by specifying the log likelihood function (2.10) and providing the starting estimates for β, a nonlinear minimizer, *nlmin*, was used to minimize $-l(y,\beta)$, and then to obtain the estimates $\hat{\beta}$ which maximized the log likelihood function $l(y,\beta)$. The nonlinear minimizer *nlmin* was based on a general quasi-Newton method. The detailed computation steps can be found in Dennis and Mei (1979) and Dennis *et al.* (1981). Different starting values were given to ensure the converge point was global.

A similar approach could be used to fit other mortality models with variable interval length. For example, to fit the generalized probit model of individual tree mortality,

(2.11) $$P=[\Phi(\beta_0+\beta_1 x_1+\cdots+\beta_k x_k)]^L,$$

where Φ is the normal cumulative distribution function, the corresponding log likelihood function is given as:

(2.12)

$$l(y,\pi(\boldsymbol{x})) = \sum_{\substack{i=1 \\ y_i=1}}^{n} L_i \log\left(\Phi\left(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}\right)\right) + \sum_{\substack{i=1 \\ y_i=0}}^{n} \log\left(1 - \Phi\left(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}\right)^{L_i}\right)$$

Using the quasi-Newton algorithm to maximize function (2.12), the maximum likelihood estimates for model (2.11) can be obtained.

Hamilton (1974) used a re-weighted least squares method to obtain the weighted least squares estimates for model (2.2) in the RISK computer program. We used the *nlmin* in Splus instead of RISK, because the *nlmin* in Splus was more accessible for us and it is easy to use, and we were not clear that the RISK was applicable to maximize the log likelihood function (2.10). With suggestions from Hamilton (1995), our further study showed that the weighted least squares estimates of model (2.2) are equivalent to the maximum likelihood estimates, if the weighting factor is equal to $\sqrt{1/\hat{p}(1-\hat{p})}$, where $\hat{p}$ is the predicted p. The weighted least squares estimation must be iterative because the weight depends on the fitted value, for which only current estimate is available. We concluded that the maximum likelihood estimates of model (2.2) from using *nlmin* and the weighted least squares estimates based on the re-weighted least squares method should be identical except for computational errors.

Besides the maximum likelihood method and the re-weighted least squares method, discriminant analysis and the weighted linear regression were also found in the literature to fit the logistic model (2.1). The later two methods are computationally quicker, but have more requirements on the data. They were not preferred to fit the model (2.2) based on the following discussions.

The discriminant function approach requires that the explanatory variable, $\boldsymbol{x}$, follows a normal distribution for two groups y=1 and y=0 having different mean and the same variance (Cornfield 1962). If there are any dichotomous explanatory variables, the discriminant function estimates will overestimate the magnitude of the association (Hosmer and Lemeshow 1989). We

did not do any implementation of the discriminant analysis to the generalized logistic model (2.2) because the growth interval L was scheduled close to 2, 5, 10, 20, and 25 years, which was not a normal distribution.

If we have multiple observations on each x and we are able to get an estimate of p which is not zero or 1 for most values of x, the weighted linear least squares can be also used to fit the logistic model (2.1) (Neter, Wasserman and Kutner 1989, page 584). When the number of observations gets large, this weighted linear regression approach to the analysis of categorical data is asymptotically equivalent to the maximum likelihood method (Hosmer and Lemeshow 1989). In the preliminary analysis for this study, we attempted the weighted linear regression approach to fit the generalized logistic model (2.2) by classifying the continuous explanatory variables into groups. Each group was treated as multiple observations on the same x, so the annual survival probability was estimated and the weighted linear regression (Neter, Wasserman and Kutner 1989, page 584) was used to estimate the parameters for the model (2.2). This approach provided estimates very similar to the maximum likelihood estimates. Moreover, the classification was useful to graphically portray the relationships between mortality and the variables. However two problems existed in this approach. First, the classification reduced the variation of the original data; Second, in some groups, the estimated survival probabilities were 1, especially for white spruce, the long-lived species. The weighted linear regression did not work in this case. Hosmer and Lemeshow (1989) suggested that the classification could be collapsed and the classes could be pooled over to eliminate the cells with $p = 1, \text{or } 0$. However, the subjective grouping of the original data is not recommended if the maximum likelihood estimation is available to solve the problem.

In summary, the maximum likelihood estimation was used to fit model (2.2). It is equivalent to the re-weighted least squares estimation, and better than the discriminant analysis and the weighted linear regression to this problem.

2.3.3 Variable selection

A mortality model that is biologically illogical cannot necessarily be expected to perform well outside the data range (Hamilton 1986). Thus an appropriate biological interpretation of the mortality process is essential in considering variable selection. In addition, statistical tests provide a useful tool when the relationships of some variables are not known. In this study, both biologically meaningful selection and statistically significant selection were considered. Information obtained from the statistical significance test was combined with an understanding of the process of mortality to select a proper set of variables. In the model fitting, the parameter significance, or the importance of the corresponding variable, was assessed through the likelihood ratio test (Hosmer and Lemeshow 1989). For example, for the j-th variable with corresponding parameter β_j, the null hypothesis is, H_0: $\beta_j = 0$ versus H_1: $\beta_j \neq 0$, and the likelihood ratio statistic is

$$\Lambda = -2\log\left(\frac{L\left(y,\left(\tilde{\beta}_0,\tilde{\beta}_1,\ldots,\tilde{\beta}_{j-1},0,\tilde{\beta}_{j+1},\ldots\tilde{\beta}_k\right)'\right)}{L\left(y,\left(\hat{\beta}_0,\hat{\beta}_1,\ldots\hat{\beta}_{j-1},\hat{\beta}_j,\hat{\beta}_{j+1},\ldots,\hat{\beta}_k\right)'\right)}\right) \tag{2.13}$$

where $(\tilde{\beta}_0,\tilde{\beta}_1,\ldots,\tilde{\beta}_{j-1},0,\tilde{\beta}_{j+1},\ldots\tilde{\beta}_k)$ are the maximum likelihood estimates under H_0, and $\hat{\beta} = (\hat{\beta}_0,\hat{\beta}_1,\ldots,\hat{\beta}_k)$ are the maximum likelihood estimates for the full model. The likelihood ratio statistic Λ compares the likelihood without variable β_j with the likelihood with variable β_j. In a large sample, Λ is approximately distributed as a chi-square distribution with degree-of-freedom 1 under H_0. A large Λ value will provide evidence to against H_0.

2.3.4 Hosmer-Lemeshow goodness-of-fit test

The Hosmer-Lemeshow goodness-of-fit test (Hosmer and Lemeshow 1989, page 140) was used to evaluate the fitting (2). In the Hosmer-Lemeshow test, first, observations were sorted in an increasing order of their estimated survival probability, then the observations were divided into g groups. As in the procedure PROC LOGISTC on SAS (SAS Institute Inc. 1992), ten groups were used in this study. The Hosmer-Lemeshow statistic was obtained by calculating the Pearson chi-square statistic from the $2\times g$ table of observed and expected frequencies (Hosmer and Lemeshow, 1989):

$$x_{HW}^2 = \sum_{i=1}^{g} \frac{\left(o_i - n_i \overline{p}_i\right)^2}{n_i \overline{p}_i \left(1 - \overline{p}_i\right)} \tag{2.14}$$

where n_i was the number of observations in the i-th group, o_i was the number of trees which were still alive at the end of the growth interval in the i-th group, and $\overline{p}_i$ was the averaged predicted survival probability for the i-th group. Under the assumption that no significant difference exists between the actual and predicted survival and mortality, the Hosmer-Lemeshow statistic should be approximately distributed as a chi-square distribution with degree-of-freedom $g - 2 = 8$.

2.4 Results

From the PSP data, 9797 white spruce, 4485 aspen and 14952 pine were used to fit model (2.2). The summary statistics for the data by live and dead trees are listed in Table 2-1 for each species. The mean diameter is around 20 cm, the stand basal area is between 15 m^2 and 70 m^2.

As outlined in the Materials section, *DBH*, *DIN*, *BA*, SC_{SP}, and *SPI* were identified as the important variables. A preliminary summary of the data showed that trees with very large *DBH* were subject to greater mortality, thus DBH^2 and DBH^2/BA were also considered for the variable selection. A negative quadratic term DBH^2 was expected to lessen the survival when *DBH* was very large, and a negative DBH^2/BA was anticipated to reduce the survival for very large trees in

old and declining stands with small *BA*. To represent the interaction between aspen and white spruce in the mixedwood stands, white spruce composition was considered as a variable in the aspen model and the aspen composition was considered in the spruce model (MacDonald 1992). A number of interaction terms based on *DBH*, *DIN*, *BA*, SC_{SP}, and *SPI* were also considered to obtain a good fit and rational interpretations. Based on the likelihood ratio test and the criteria that a rational interpretation is more important than the model precision, the significant explanatory variables for the three species were:

White spruce	DBH, DBH^2, DIN, SPI, DBH^2/BA
Aspen	DBH, DBH^2, DIN, SC_{SW}, SPI/BA, DBH^2/BA
Lodgepole pine	DBH, DBH^2, DIN, SC_{PL}, SPI, DBH^2/BA

where SC_{SW} is white spruce composition, and SC_{PL} is lodgepole pine composition. The significant variables and interaction terms were different for the three species, which indicates that the causes of mortality for the three species are not identical, and to some extent the relations between mortality and explanatory variables differ considerably.

The corresponding models were:

white spruce model (2.15)

$$P = \left[1 + \exp\left(-\left(\beta_0 + \beta_1 DBH + \beta_2 DBH^2 + \beta_3 DIN + \beta_4 SPI + \beta_5 \frac{DBH^2}{BA}\right)\right)\right]^{-L},$$

aspen model (2.16)

$$P = \left[1 + \exp\left(-\left(\beta_0 + \beta_1 DBH + \beta_2 DBH^2 + \beta_3 DIN + \beta_4 SC_{sw} + \beta_5 \frac{SPI}{BA} + \beta_6 \frac{DBH^2}{BA}\right)\right)\right]^{-L},$$

and lodgepole pine model (2.17)

$$P = \left[1 + \exp\left(-\left(\beta_0 + \beta_1 DBH + \beta_2 DBH^2 + \beta_3 DIN + \beta_4 BA + \beta_5 SC_{pl} + \beta_6 SPI + \beta_7 \frac{DBH^2}{BA}\right)\right)\right]^{-L}.$$

The fit statistics for the above three models are listed in Tables 2-2, 2-3 and 2-4 for white spruce, aspen, and pine. All parameters are significant. The Hosmer-Lemeshow test was used to test the goodness-of-fit. The observed live tree number and predicted live tree number for each group are also listed in Tables 2-2, 2-3, and 2-4. At level $\alpha = 0.05$, there was no significant difference between the predicted and actual survival probability.

An example of the model application is given here. For a white spruce with diameter (*DBH*) 2 cm and annual diameter increment (*DIN*) 0.2 cm in a stand with basal area (*BA*) 20 m^2 per ha and site productivity index (*SPI*) 18 m, based on the fitted generalized logistic model, the estimate of the 5-year survival probability is equal to

$$P = \left[1 + \exp\left[-\left(5.578705 + 0.067434\,DBH - 0.000852717\,DBH^2 + 1.126298\,DIN - 0.00763\,BA - 0.096\,SPI - 0.012327\,\frac{DBH^2}{BA}\right)\right]\right]^{-5}$$

$$= 0.917449.$$

Sometimes *DIN* is not available in a forest inventory, and only one-time measurements of trees and stands are available to predict the survival probabilities. In that case, a simpler mortality model is needed. Following the same methodology, a mortality model without *DIN* as an explanatory variable was also fitted to the PSP data set. The fitting statistics are attached in Tables 2-5, 2-6, and 2-7 for white spruce, aspen, and lodgepole pine. The Hosmer-Lemeshow goodness-of-fit test failed for all three models, which indicates that the diameter increment is a very important input variable to project the survival probability precisely. However, from the predicted and actual number of live trees in each group listed in Tables 2-5, 2-6, and 2-7, it can be seen that the differences are not great, and they may be acceptable for applications.

2.5 Discussion

The generalized logistic models of individual tree mortality for white spruce (2.15), aspen (2.16), and lodgepole pine (2.17) confirm some commonly held theories about Alberta mixedwood stands that are older than 30 years. Figures 2-1 to 2-3 display the fitted survival surface for the three species under different conditions of diameter (*DBH*), diameter increment (*DIN*), total basal area (*BA*), species composition (SC_{SP}), and site productivity index (*SPI*). Each figure has 4 charts. The first chart shows the impact of *DIN* when other stand characters are fixed, and the remaining 3 charts show the impacts of *BA*, SC_{SP}, and *SPI* respectively when trees are growing at 0.15 cm in diameter annually, i.e., *DIN*=0.15. White spruce is the longest lived species with the highest survival probability (Figure 2-2), while aspen is the shortest lived species with lowest survival probability (Figure 2-1). Even with large stand basal area (*BA*) and aspen composition (SC_{AW}), the shade tolerant spruce is still able to maintain a high survival. Aspen survival probability is more sensitive to the effects of stand basal area, species composition, and site productivity. Pine survival is generally lower than spruce and higher than aspen (Figure 2-3).

For all three models, the positive coefficients of *DBH* and negative coefficients of DBH^2 indicate that aspen, spruce, and pine trees have relatively lower survival when they are small in size. With increasing *DBH* the survival probability becomes larger, stays relatively stable, and then decreases. This trend can be seen clearly in Figures 2-1, 2-2, and 2-3. In young and dense stands, significant mortality of small trees is expected due to self-thinning. In older stands smaller trees are likely suppressed by larger trees, resulting in a lower survival probability. In Figures 2-1(b), 2-2(b), and 2-3(b), it is evident that increasing total basal area reduces small tree survival. The total basal area of a stand indicates the magnitude of competition among trees.

Median and large size trees will usually experience less competition, and thus have higher and stable survival (Figures 2-1, 2-2, and 2-3).

Figures 2-1, 2-2, and 2-3 also show that the survival probability decreases with very large *DBH*. As a tree increases in size, the photosynthetic efficiency, a function of foliar biomass, increases to a maximum value and then decreases as self-shading becomes more severe. At the same time, respiration, a function of total biomass, increases as long as the tree is growing. As net photosynthesis and respiration converge, tree growth slows significantly and maximum size is reached. The poor carbon balance of the tree leads to an inherent decline in vigour. Probability of mortality at this stage increases dramatically as a result of loss of vigour and reduced ability to survive damages by diseases, insects, climatic extremes, or other external factors (MacDonald 1994).

Trees with very large *DBH* have lower survival when the stand basal area is small, especially for aspen and pine (Figures 2-1(b), and 2-3(b)). This is likely indicative of very old stands. When trees have reached full size and mortality increases, gaps occur in the canopy, thereby subjecting the stand to increased exposure to wind, and sunlight. Large, old aspen and pine appear to be intolerant of these stresses, and become vulnerable to windthrow, disease and insect attack (Peterson and Peterson, 1992).

Trees with large diameter increments have higher survival probabilities (Figures 2-1(a), 2-2(a), and 2-3(a)). With decreasing the diameter increment, aspen survival declines more dramatically than spruce and pine, which indicates that the intolerant aspen is more sensitive to life stresses.

In aspen and spruce mixedwood stands, the shade intolerant aspen shows faster early growth and rapidly dominates the site by occupying the upper layer of the canopy (Peterson and Peterson, 1992). Therefore aspen has the competitive advantage over shade tolerant white spruce which exhibits slow juvenile growth. For about 50 years, aspen continues to dominate the site,

but white spruce becomes progressively more conspicuous in the understorey. The longer-lived white spruce eventually dominates as the short-lived aspen begins to die out (Peterson and Peterson 1992). In the model for aspen it is seen that, with greater spruce composition, there is less survival of aspen (Figure 2-1(c)). High spruce composition suggests that the site may not be suitable for aspen growth or that spruce has begun to replace aspen in the overstorey. In the later case, aspen would generally be close to natural mortality age and thus have lower survival. Because of their shade intolerance younger aspen cannot survive when overtopped by white spruce. On the other hand, in the spruce model, aspen composition does not have a significant influence on spruce mortality (Figure 2-2(c)). This indicates that spruce, a long-lived and shade tolerant species, can survive under an aspen canopy for a long time.

Pine survival is greater in pure than in mixedwood stands (Figure 2-3(c)). With decreasing pine composition, the survival at a given *DBH* declines, particularly with very small *DBH*. As for aspen, it could be that sites with low pine composition are not suited to pine growth. Since pine is shade intolerant, small trees are very likely to be suppressed by larger trees and so have low survival. The mortality of large pines in these mixedwood stands likely relates to death of the few large old remnant pine individuals in a stand which is succeeding to a different species composition.

The negative sign of *SPI* in the pine model indicates that increasing site productivity increases the mortality of pine (Figure 2-3(d)). This is especially true for small trees. A Similar trend was shown in Crow and Hick (1990) for mixed oak stands. On better sites, trees grow faster, so that competition may be more severe, increasing the probability of death. This is the reason that trees have lower survival probability on richer sites, especially the small trees that experience more competition. In the stand growth model build by Tait *et al.* (1988) for lodgepole pine in Alberta, the negative impact of high site productivity on survival was also shown.

The higher mortality of spruce on richer sites (Figure 2-2(d)) may also be due to more severe competition from spruce, and also from aspen which survives better on richer sites (Figure 2-1(d)). Since white spruce is more shade tolerant than pine, the competition impact on spruce (Figure 2-2(d)) is not as dramatic as on pine (Figure 2-3(d)). In the aspen model, mortality is slightly higher on poor sites (Figure 2-1(d)), which is opposite to pine and spruce. Although less competition is expected on poor sites, aspen may have higher mortality due to water and nutrients limitations.

In summary, the generalized logistic model was fitted for individual tree mortality for aspen, white spruce, and lodgepole pine in Alberta boreal mixedwood stands. The fitted model for each species describes the natural phenomenon of mortality in mixedwood stands, and supports a logical biological interpretation of relationships between the survival probability and the explanatory variables. Since PSP data only include stands older than 30 years, the model should not be used in juvenile stands.

Table 2-1. Summary statistics for tree and stand characteristics by life condition code

Species	Life condition	N	Variable	Mean	Std Dev	Minimum	Maximum
Spruce	Live	8940	*DBH*	21.919564	13.493280	1.100000	77.500000
			DIN	0.166055	0.161000	-3.040000	1.080000
			BA	41.040697	9.055433	15.521100	67.207900
			SPI	16.232507	2.233793	9.640000	21.010000
			SC_{AW}	0.171163	0.218552	0.000000	1.000000
			L	8.254530	3.017931	2.200000	14.700000
	Dead	857	*DBH*	19.154959	14.698392	1.100000	65.800000
			DIN	0.075204	0.181866	-3.650000	0.950000
			BA	42.310776	8.675435	18.899500	62.213500
			SPI	16.527095	1.921749	10.030000	21.010000
			SC_{AW}	0.1598600	0.210748	0.000000	0.900000
			L	8.878063	2.730805	2.200000	14.700000
Aspen	Live	3649	*DBH*	25.976788	11.006783	1.000000	69.300000
			DIN	0.173990	0.198761	-5.270000	1.340000
			BA	38.289816	9.948048	14.949800	62.406300
			SPI	16.728838	1.909742	11.540000	20.840000
			SC_{SW}	0.246067	0.239131	0.000000	1.000000
			L	8.904878	3.278919	2.200000	14.700000
	Dead	859	*DBH*	22.030151	12.217231	1.100000	68.100000
			DIN	0.093527	0.171612	-2.890000	0.700000
			BA	37.998905	9.094311	15.650900	62.213500
			SPI	16.648557	1.800844	11.540000	20.840000
			SC_{SW}	0.264933	0.243146	0.000000	1.000000
			L	9.425611	2.902887	2.200000	14.700000
Pine	Live	12896	*DBH*	17.416711	7.349325	1.000000	57.400000
			DIN	0.126999	0.110327	-2.620000	1.260000
			BA	37.915107	8.728114	15.521100	62.213500
			SPI	17.298752	2.148105	9.640000	21.010000
			SC_{PL}	0.856754	0.206071	0.000100	1.000000
			L	10.209716	2.805880	2.200000	14.700000
	Dead	2056	*DBH*	12.095185	6.235694	1.000000	45.000000
			DIN	0.037417	0.074377	-1.000000	0.570000
			BA	37.979406	8.303816	13.485700	62.213500
			SPI	17.562461	1.924737	9.640000	21.010000
			SC_{PL}	0.865856	0.201868	0.000100	1.000000
			L	10.743093	2.300858	2.200000	14.700000

Table 2-2. Parameter estimates for white spruce mortality model (2.15)

Variable	Degree-of-freedom	Parameter	Parameter estimate
INTERCEPT	1	β_0	5.578705
DBH	1	β_1	0.067434
DBH^2	1	β_2	-0.000852717
DIN	1	β_3	1.126298
BA	1	β_4	-0.007630
SPI	1	β_5	-0.096000
DBH^2/BA	1	β_6	-0.012327

Log likelihood function is -2773.807

Log likelihood function ($\beta_0=\beta_1=\beta_2=\beta_3=\beta_4=\beta_5=\beta_6=0$) = -51159.07

Hosmer and Lemeshow goodness-of-fit test

		Number of live trees	
Group	Total number of trees	Observed	Expected
1	980	779	787.0315
2	980	852	856.4415
3	980	867	877.6153
4	980	886	890.3504
5	980	886	900.0443
6	980	927	908.9999
7	980	921	917.5572
8	980	942	926.7516
9	980	937	936.3679
10	977	943	945.3230

Hosmer-Lemeshow statistic = 14.67209 with 8 DF (p-value=0.06584798)

Table 2-3. Parameter estimates for aspen mortality model (2.16)

Variable	Degree-of-freedom	Parameter	Parameter estimate
INTERCEPT	1	β_0	1.716708
DBH	1	β_1	0.098991
DBH^2	1	β_2	-0.00117481
DIN	1	β_3	7.490443
SC_{SW}	1	β_4	-0.696387
DBH^2/BA	1	β_5	-0.022964
SPI / BA	1	β_6	0.097591

Log likelihood function is -1901.778

Log likelihood function ($\beta_0=\beta_1=\beta_2=\beta_3=\beta_4=\beta_5=\beta_6=0$) = -22413.65

Hosmer and Lemeshow goodness-of-fit test

		Number of live trees	
Group	Total number of trees	Observed	Expected
1	449	207	215.3984
2	449	286	303.5154
3	449	335	337.4708
4	449	355	358.0172
5	449	376	373.7513
6	449	394	387.0221
7	449	404	398.6377
8	449	419	410.1008
9	449	429	421.8733
10	444	424	431.8216

Hosmer-Lemeshow statistic = 14.96926 with 8 DF (p-value=0.05974602)

Table 2-4. Parameter estimates for lodgepole pine mortality model (2.17)

Variable	Degree-of-freedom	Parameter	Parameter estimate
INTERCEPT	1	β_0	3.824287
DBH	1	β_1	0.253224
DBH^2	1	β_2	-0.00225115
DIN	1	β_3	0.810339
BA	1	β_4	-0.002153
SC_{PL}	1	β_5	1.324298
SPI	1	β_6	-0.198713
DBH^2/BA	1	β_7	-0.049673

Log likelihood function is -5133.691

Log likelihood function ($\beta_0=\beta_1=\beta_2=\beta_3=\beta_4=\beta_5=\beta_6=0$) = -91268.82

Hosmer and Lemeshow goodness-of-fit test

		Number of live trees	
Group	Total number of trees	Observed	Expected
1	1495	849	890.4359
2	1495	1074	1121.689
3	1495	1218	1219.743
4	1495	1317	1281.818
5	1495	1346	1326.811
6	1495	1366	1362.682
7	1495	1421	1392.334
8	1495	1419	1416.91
9	1495	1430	1437.637
10	1497	1456	1464.807

Hosmer-Lemeshow statistic = 15.10788 with 8 DF (p-value=0.55000)

Table 2-5. Parameter estimates for the white spruce mortality model without considering *DIN*

Variable	Degree-of-freedom	Parameter	Parameter estimate
INTERCEPT	1	β_0	5.671267
DBH	1	β_1	0.089780
DBH^2	1	β_2	-0.00121954
BA	1	β_3	-0.010316
SPI	1	β_4	-0.102499
DBH^2/BA	1	β_5	-0.009349

Log likelihood function is -2810.329

Log likelihood function ($\beta_0=\beta_1=\beta_2=\beta_3=\beta_4=\beta_5=\beta_6=0$) = -51159.07

Hosmer and Lemeshow goodness-of-fit test

		Number of live trees	
Group	Total number of trees	Observed	Expected
1	980	796	787.3876
2	980	876	852.2263
3	980	879	875.0859
4	980	880	889.7681
5	980	890	900.1622
6	980	893	909.228
7	980	932	917.9677
8	980	929	927.0753
9	980	938	936.3322
10	977	927	945.8126

Hosmer-Lemeshow statistic = 17.56454 with 8 DF (p-value=0.000564542)

Table 2-6. Parameter estimates for the aspen mortality model without considering *DIN*

Variable	Degree-of-freedom	Parameter	Parameter estimate
INTERCEPT	1	β_0	1.703197
DBH	1	β_1	0.111998
DBH^2	1	β_2	-0.00109154
SC_{sw}	1	β_3	-0.738333
DBH^2/BA	1	β_4	-0.015611
SPI / BA	1	β_5	1.308282

Log likelihood function is -2100.314

Log likelihood function ($\beta_0=\beta_1=\beta_2=\beta_3=\beta_4=\beta_5=\beta_6=0$) = -22530.11

Hosmer and Lemeshow goodness of fit test for the model above

		Number of live trees	
Group	Total number of trees	Observed	Expected
1	450	248	265.8279
2	450	340	329.1953
3	450	374	349.4069
4	450	358	361.4055
5	450	387	370.4001
6	450	379	378.4422
7	450	383	385.2188
8	450	381	392.7138
9	450	390	404.2797
10	458	409	418.9188

Hosmer-Lemeshow statistic = 27.40428 with 8 DF (p-value=0.0006018415)

Table 2-7. Parameter estimates for the lodgepole pine mortality model without considering *DIN*

Variable	Degree-of-freedom	Parameter	Parameter estimate
INTERCEPT	1	β_0	3.786151
DBH	1	β_1	0.293849
DBH^2	1	β_2	-0.00325343
BA	1	β_3	-0.00055858
SC_{PL}	1	β_4	1.373082
SPI	1	β_5	-0.218494
DBH^2/BA	1	β_6	-0.044720

Log likelihood function = -5155.49.

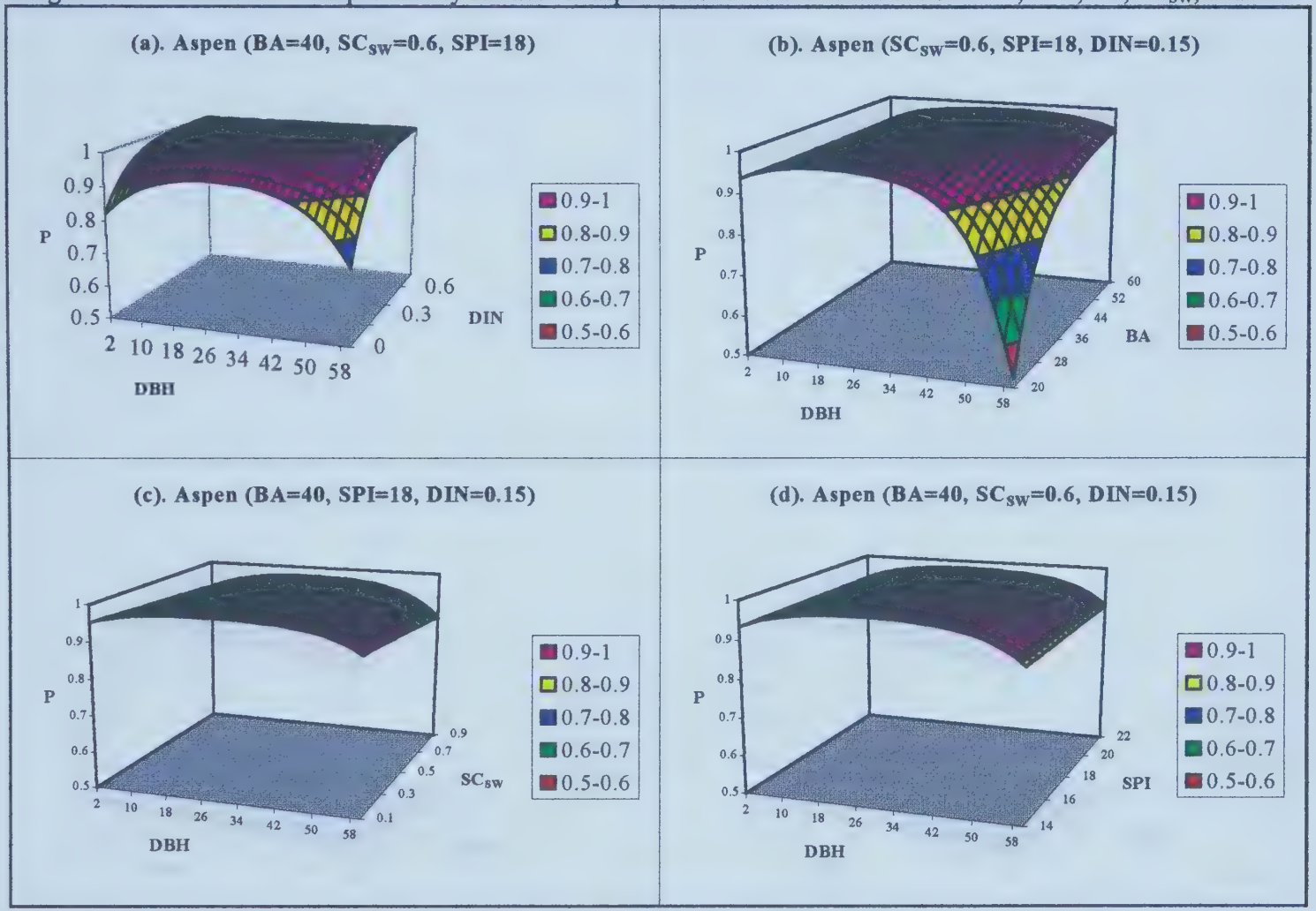
Log likelihood function ($\beta_0=\beta_1=\beta_2=\beta_3=\beta_4=\beta_5=\beta_6=0$) = -91268.82

Hosmer and Lemeshow goodness-of-fit test

		Number of live trees	
Group	Total number of trees	Observed	Expected
1	1495	826	865.2778
2	1495	1093	1115.1150
3	1495	1215	1219.9987
4	1495	1324	1284.5533
5	1495	1347	1330.3012
6	1495	1387	1366.5032
7	1495	1415	1396.1953
8	1495	1420	1420.0891
9	1495	1424	1439.6116
10	1497	1445	1466.0804

Hosmer-Lemeshow statistic = 43.23318 with 8 DF (p-value=0.00000008)

Figure 2-1. The fitted survival probability surface for aspen under different conditions of *DBH*, *DIN*, *BA*, SC_{SW}, and *SPI*



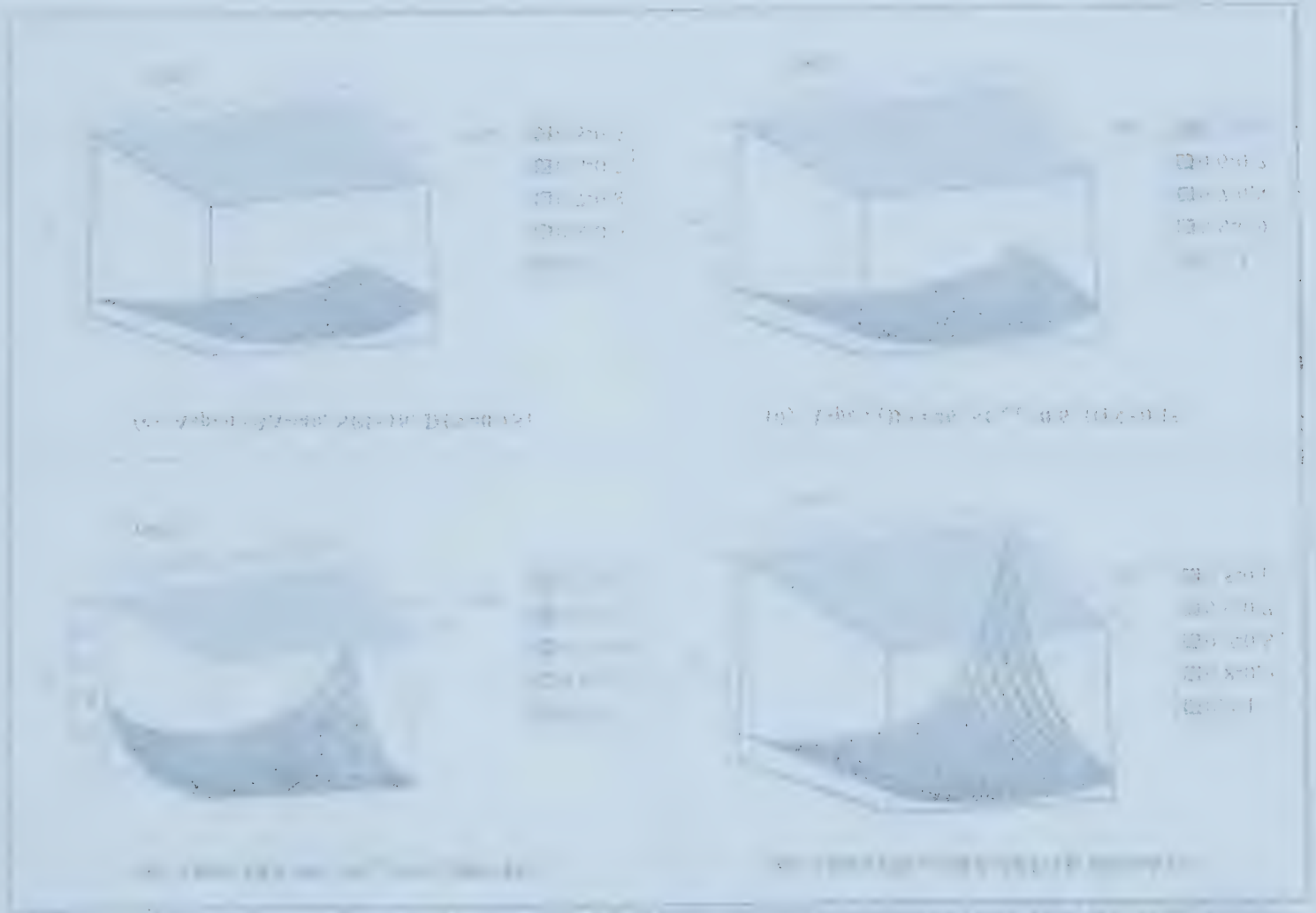
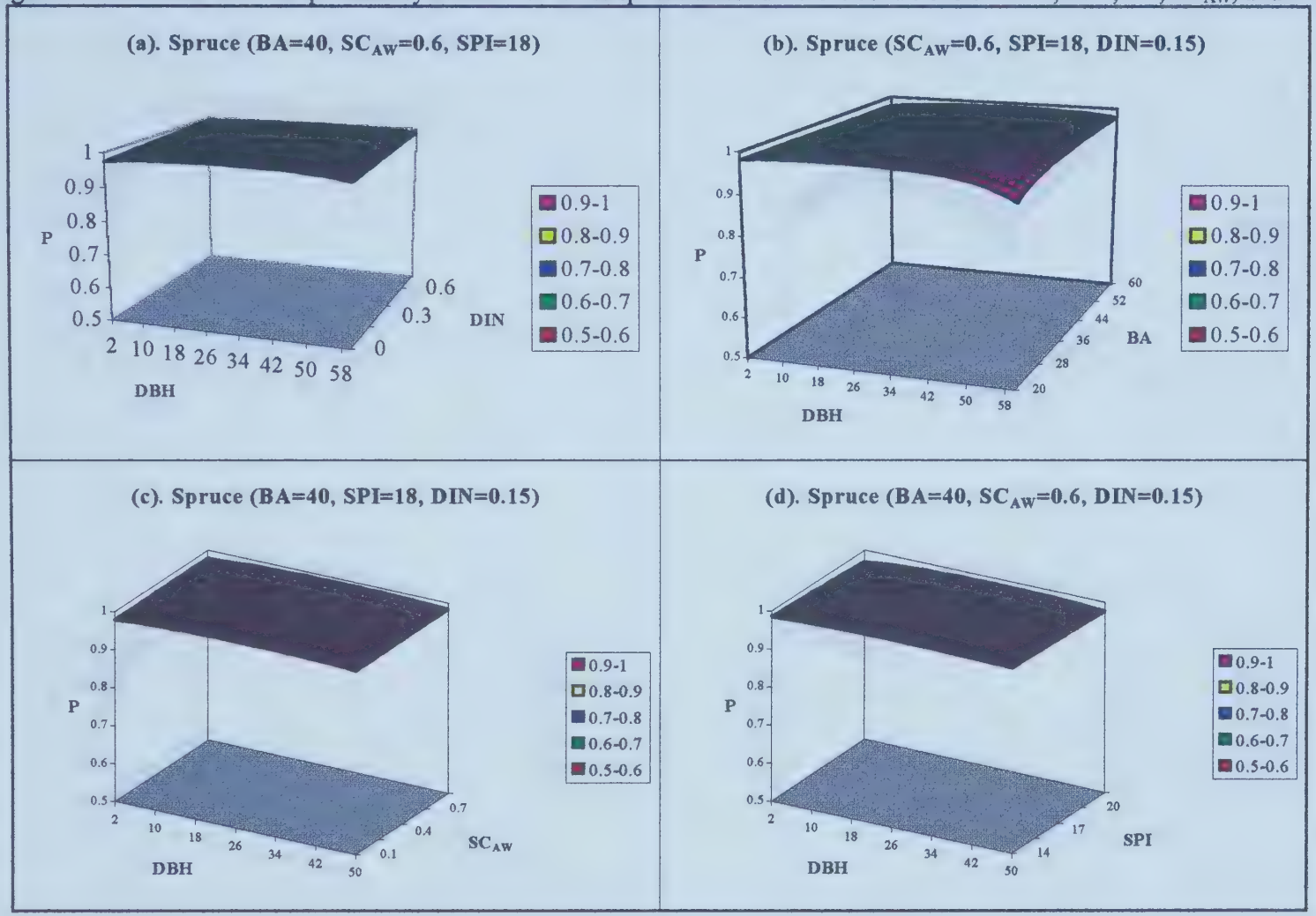


Figure 2-2. The fitted survival probability surface for white spruce under different conditions of *DBH*, *DIN*, *BA*, SC_{AW}, and *SPI*



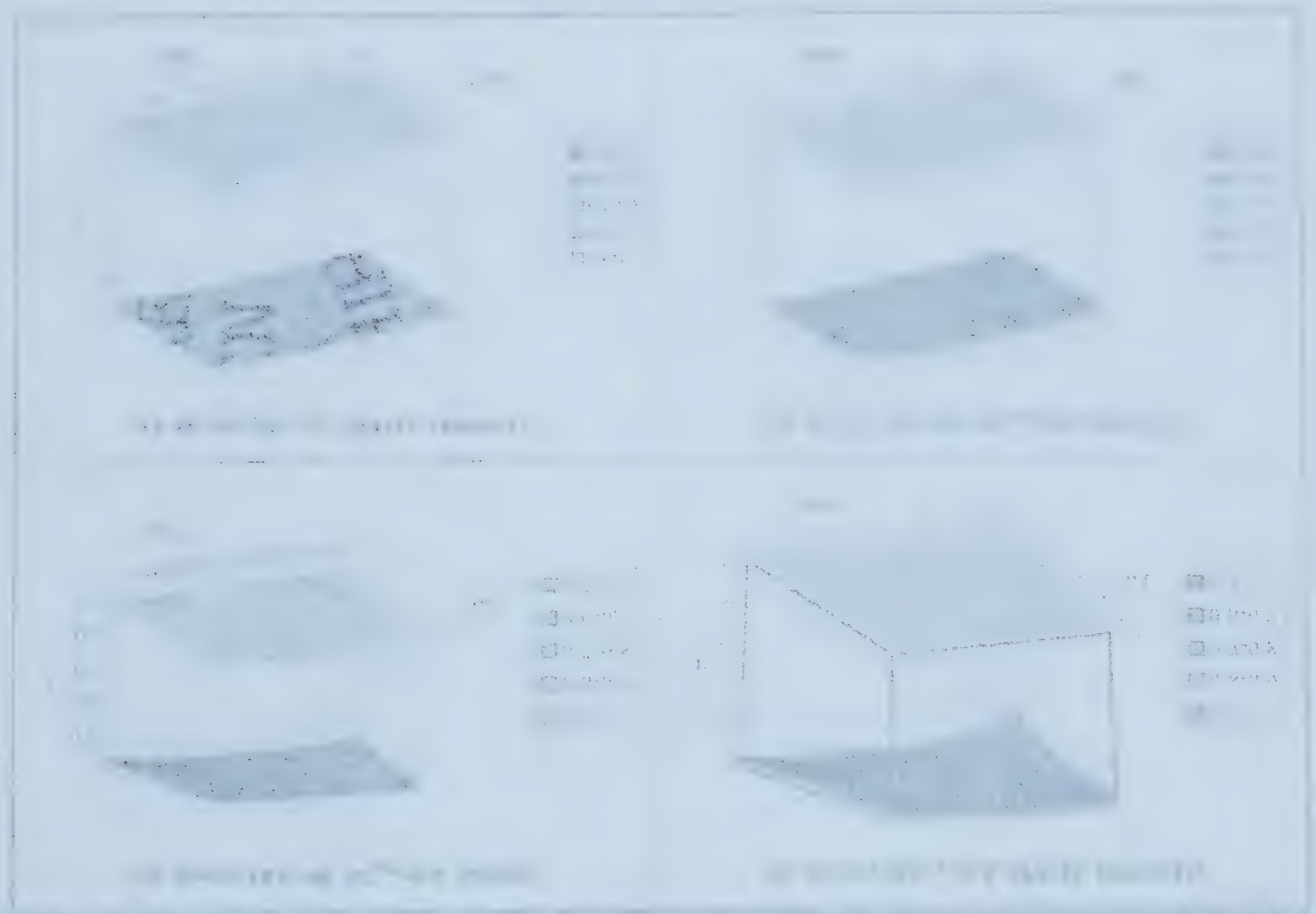
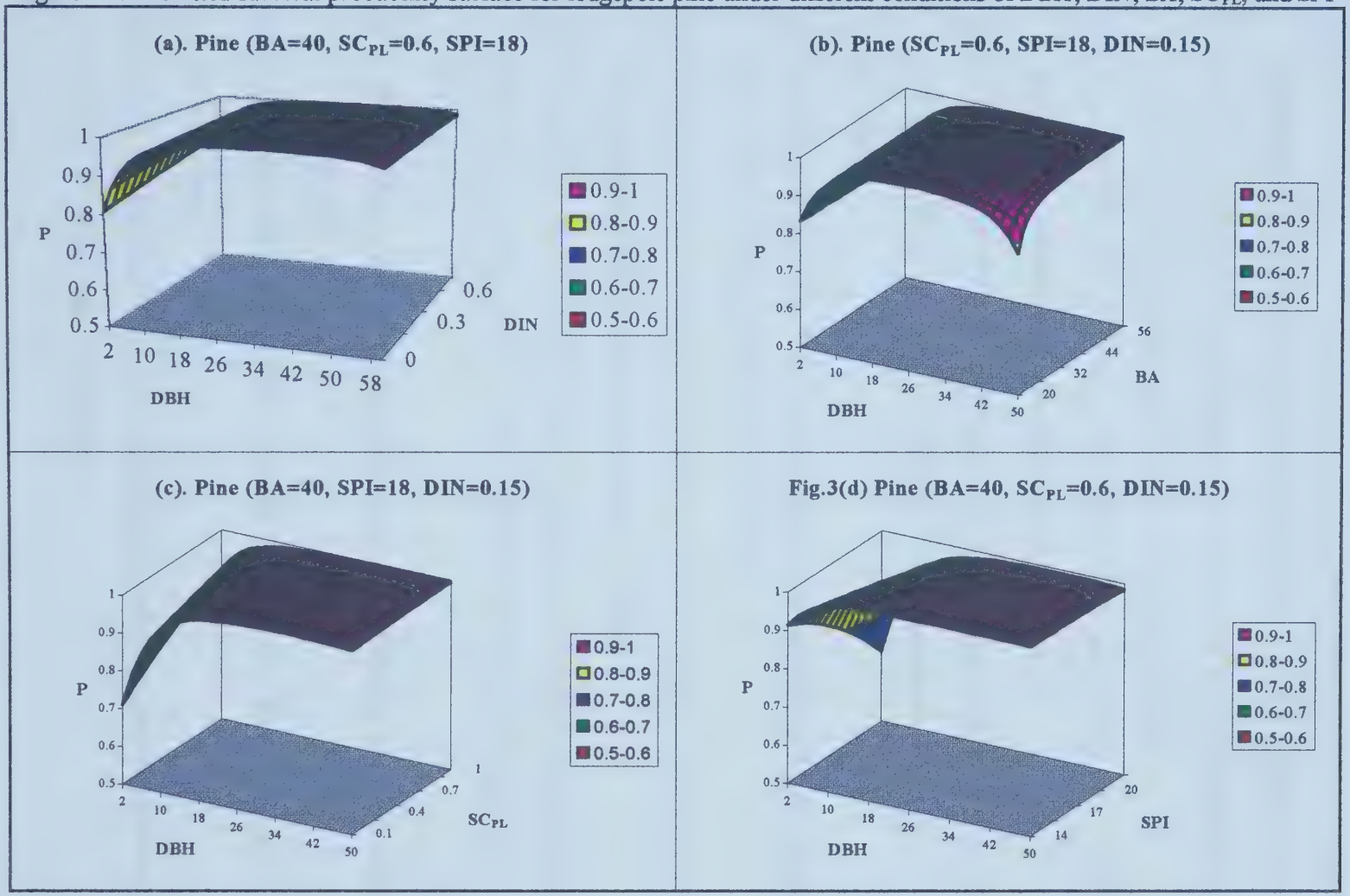


Figure 2-3. The fitted survival probability surface for lodgepole pine under different conditions of *DBH*, *DIN*, *BA*, SC_{PL}, and *SPI*



2.6 References

Alberta Forest Service. 1990. Permanent sample plots: field procedure manual. Timber Management Branch, Alberta Forest Service, Edmonton, Alberta. FMOPC 83-03.

Buchman, R.G., Pederson, S.P., and Walters, N.R. 1983. A tree survival model with application to species of the Great Lakes region. Can. J. For. Res., 13:601-608.

Clutter, J.L., Fortson, J.C., Pienaar, L.V., Brister, G.H., and Bailey, R.L. 1983. Timber management: a quantitative approach. John wiley & sons.

Cornfield, J. 1962. Joint dependence of the risk of coronary heart disease on serum cholesterol and systolic blood pressure: A discriminant function analysis. Federation proceedings, 21, 58-61.

Crow, G.R. and Hicks, R.R. Jr. 1990. Predicting mortality in mixed oak stands following spring insect dafoliation. Forest Science, 36 (3): 831-841.

Dennis, J.E., Gay, D.M., and Welsch, R.E. 1981. An adaptive nonlinear least-squares algorithm. ACM Transactions on Mathematical Software. 7: 348-383.

Dennis, J.E. and Mei, H.H.W. 1979. Two new unconstrained optimization algorithms which use function and gradient values. Journal of Optimization Theory and Applications. 28:453-483.

Ek, A.R. 1974. Nonlinear models for stand table projection in Northern hardwood stands. Can. J. For. Res., 4:23-27.

Hamilton, D.A. 1974. Event probabilities estimated by regression. USDA For. Serv. Res. Pap. INT-152, 18p.

Hamilton, D.A. and Edwards, B.M. 1976. Modeling the probability of individual tree mortality. U.S.D.A. For. Serv. Res. Pap. INT-185.

Hamilton, D.A. 1986. A logistic model of mortality in thinned and unthinned mixed conifer stands of Northern Idaho. Forest Science, 32:989-1000.

Hamilton, D.A. 1990. Extending the range of applicability of an individual tree mortality model. Can. J. For. Res., 20:1212-1218.

Hamilton, D.A. 1995. Personal Communication. USDA Forest Service, Intermountain Research Station, Moscow, Idaho.

Hann, D.W. 1980. Development and evaluation of an even- and uneven-aged ponderosa pine / Arizona fescue stand simulator. U.S.D.A. For. Serv. Res. Pap. INT-267.

Harms, W.R. 1983. An empirical function for predicting survival over a wide range of densities. Proc. Second Biennial South Silv. Res. Conf., Atlanta, Ga, Nov 4-5, 1982. USDA Forest Serv. Gen.Tech. Rep. SE-24, p 334-337.

Hinkley, D.V., Reid, N., and Snell, J. 1991. Statistical theory and modeling. Chapman and Hall Ltd.

Hosmer, D.W. and Lemeshow, S. 1989. Applied logistic regression. John Wiley & Sons. New York. USA.

Huang, S. 1992. Diameter and height growth models. Ph.D. Thesis, Department of Forest Science, University of Alberta.

Huang, S, and Titus, S.J. 1993. An index of site productivity for uneven-aged and mixed-species stands. Can. J. For. Res., 23:558-562.

Lee, Y.J. 1971. Predicting mortality for even-aged stands of lodgepole pine. The For. Chron. 47:29-32.

MacDonald, S.E. 1992. Personal communication. Department of Renewable Resources, University of Alberta.

MacDonald, S.E. 1994. Personal communication. Department of Renewable Resources, University of Alberta.

McCullagh, P., and Nelder, J.A. 1983. Generalized linear models. Chapman and Hall Ltd.

Monserud, R.A. 1976. Simulation of forest tree mortality. Forest Science, 22:438-444.

Moser, J.W. 1972. Dynamics of an uneven-aged forest stand. Forest Science, 18:184-191.

Neter, J., Wasserman, W., and Kutner, M.H. 1989. Applied linear regression models. 2nd Edition. Homewood Ill. : R.D.. Irwin.

Peterson E.B. and Peterson, N.M. 1992. Ecology, management, and use of aspen and balsam poplar in the prairie provinces. Forestry Canada.

SAS Institute Inc. 1992. SAS user's guide: statistics. Version 6 ed., Cary, NC 27512-8000.

SAS Institute Inc. 1992. SAS technical report P-229, SAS/STAT software: Changes and enhancement. Cary, NC 27513.

Somers, G.L., Oderwald, R.G., Harms, W.R., and Langdon, O.G. 1980. Predicting mortality with a Weibull distribution. Forest Science, 26:291-300.

Statistical Science. 1993. S-PLUS Guide to statistical and mathematical analysis, version 3.2. Seattle: StatSci, a division of mathsoft Inc.

Tait, D.E., Cieszewski, C.J. and Bella, I.E. 1988. The stand dynamics of lodgepole pine. Canadian Journal of Forest Research, 18:1255-1260.

Chapter 3

A logistic model of juvenile tree mortality

3.1 Introduction

Juvenile trees in regenerated stands are vulnerable to many agents, such as suppression, flooding, climatic extremes, diseases, insects, and other animal attack. One individual agent or a combination of these agents can cause sudden and extensive mortality in juvenile stands. Although many studies have discussed juvenile tree mortality ecologically and biologically (Wakeley 1954, Shiver *et al.* 1990, Turner 1990, Morse and Kulman 1984, Barbour *et al.*1990), juvenile tree mortality still remains one of the most difficult aspects to precisely and quantitatively describe and predict. For the last several decades in Alberta, many natural stands have been harvested and replaced by regenerated stands. How do these regenerated stands develop? What are the basic growth relations? Answers to these questions are needed by both industry and government agencies. To meet this need, it is imperative to build a regeneration stand growth model system. The juvenile tree mortality model is one of the essential parts of the model. This chapter presents a juvenile tree mortality model for Alberta regenerated stands.

Logistic regression has been widely used to model individual mortality in economic, biological, and medical studies (Hosmer and Lemeshow 1989, Maddala 1988). Since Hamilton (1974) and Monserud (1976) applied the logistic model to individual tree mortality, many others have also used a similar approach (Hann 1980, Buchman and Pederson 1983, Hamilton 1986, Hamilton 1990). We previously used a generalized logistic model to model the survival probability of an individual tree in Alberta mixedwood stands older than 30 years (Chapter 2). The current study is an extension to the juvenile stands. The data for this study are permanent sample plots, established in regenerated stands younger than 15 years, with remeasurements

every two years. Since the remeasurement interval had an equal length, 2 years, the modeling technique in this study is more straight forward than in Chapter 2. The logistic model was directly used to relate the survival probability of an individual to the tree and stand characteristics. The relations were evaluated for three major species: trembling aspen (*Populus tremuloides* Michx), white spruce (*Picea glauca* (Moench) Voss), and lodgepole pine (*Pinus contorta* var. *latifolia* Engelm). The maximum likelihood method was used to estimate parameters, and Hosmer-Lemeshow test was used to test the goodness-of-fit. In the fitted model, the survival probability can be predicted, given age, height, stand density, species composition, and site productivity.

3.2 Materials

3.2.1 Data

The Alberta Land and Forest Services provided the data for this study. The data are permanent sample plots designed to "monitor" juvenile stand establishment and growth on cut-blocks (Gilday 1990). The plots were established in various ecoregions across the province since 1980. There were 107 plots. The plots were remeasured every two years and with same plots remeasured up to 7 times. At plot establishment, harvest, scarification, regeneration treatment, ecoregion, moisture regime, drainage condition, duff depth and some other stand characteristics were recorded for each plot. The regeneration treatments included left-for-natural, seeded, and planted with different species. Of the 107 plots, 69 were planted with white spruce in which 33 were planted with 1-year-old containerized seedlings and 36 were planted with 3-year-old bareroot seedlings; 22 were left for natural or seeded with lodgepole pine or mixed species; 7 were planted with lodgepole pine; 3 were seeded with white spruce; and 6 were seeded with forage and unspecified species. The last three groups were not of sufficient size for analysis and 6 plots were re-seeded or thinned. They were excluded from this study.

Each plot consisted of 40 quadrats (2 m × 2.5 m) that were alternately established on each side of four 25 metre transects distributed over a cut-blocks (Gilday 1990). Lodgepole pine, jack pine, white spruce, black spruce, and all fir species were defined as acceptable species. In a quadrat, each tree of acceptable species and the three tallest deciduous trees were tagged and measured. Measurements included species, age, height, regeneration origin, damage code, and life condition code. Other deciduous trees were counted by species. Tree age was estimated when a tree was first found on the site. The tree regeneration origin indicated whether the tree was planted or seeded. The damage code indicated the damage categories for live trees and reported possible causes of mortality when trees were first found dead (Gilday 1990). It included 21 catergories such as suppression, flooding, climatic extremes, diseases, insects, and etc. The life condition code was assigned 1 if a tree was alive, and 0 if the tree was found dead.

3.2.2 Response and explanatory variables

After establishment, all the plots were remeasured every two years in summer. Each non-overlapping two-year period defined a growth interval. For each interval, the measurements of a tree at the beginning and the life condition code for the same tree at the end formulated one observation for the mortality modeling. Excluding the last measurement, all the trees have their life condition response at the end of the two-year interval. The response is a binary response variable.

The data consisted of repeated measurements on the same trees over an extended period, and it did not satisfy the assumption of independence. However, because the number of the multiple measurements for the same trees was very small in comparison with the number of trees, the serial correlation was not expected to be serious (Vanclay 1994). Also, using the whole data set won't produce a biased model, although the variance may be underestimated (Vanclay

1994). In this study, the data were used to fit the model, and the serial correlation problem was addressed after the model fitting.

To consider the explanatory variables, tree age and height were very important to indicate a tree's growth, vigor, and the ability to resist internal and external damage agents. In the data, both age and height were measured, and considered as explanatory variables.

The tree regeneration origin had three classes: *BAREROOT* for trees which were nursery grown bareroot seedlings, *CONTAINER* for trees which were nursery grown containerized seedlings, and ingress for trees which were germinated from natural or artificial seeding on the site. Since all blocks were harvested by clear-cutting, a small number of advance trees, those which had grown on the site for at least one year prior to harvesting, were excluded from this study. Two indicator variables, *BAREROOT* and *CONTAINER*, were used to indicate the tree regeneration origin. If the tree was a nursery grown bareroot seedling, then *BAREROOT*=1 and *CONTAINER*=0. If the tree was a nursery grown containerized seedling, then *BAREROOT*=0 and *CONTAINER*=1. If the tree was an ingress seedling, then *BAREROOT*=0 and *CONTAINER*=0.

The damage code included 21 catergories such as suppression, flooding, climatic extremes, deseases, insects, and etc. A preliminary study showed that the survival probability in two years were not significantly different among the categories, but different between damaged and healthy seedlings. Therefore only a damage indicator was considered. *DAMAGE*=1 if a tree was damaged, and *DAMAGE*=0 if it was healthy.

Although not all the deciduous trees were tagged and remeasured, all of them were counted by species. This made it possible to summarize total density, density by species, and species composition as additional variables. Total density (*DENSITY*) was defined as the total number of trees for all species per ha. *AWDEN*, *SWDEN*, and *PLDEN* were aspen density, spruce density, and lodgepole pine density. Species composition (SC_{SP}) was defined as the number of trees per ha for a target species divided by number of trees per ha for all species.

The site productivity classification of the cut-blocks was identified from the Phase 3 Inventory. Almost all the cut-blocks were located in good and medium sites, thus the site class had only two categories: good and medium. *SITE*=1, if the site class was good, and *SITE*=0 if the site class was medium. In addition, the drainage condition, an assessment of soil drainage characteristics of each block, and the humus depth, the average depth of the duff layer on the block, were also considered. In the original data, the drainage condition had 12 classes and humus depth had 5 classes. In this study only two classes were considered for drainage condition, well-drained and poorly-drained. *DW*=1 if the stand was well-drained; and *DW*=0 if the stand was poorly-drained. The humus depth was divided into deep (≥ 16 cm), where *HS*=0; and shallow (< 16 cm), where *HS*=1.

3.3 Methods

3.3.1 Model specification

Assume that there are *k* explanatory variables $\boldsymbol{x} = (x_1, x_2, \cdots, x_k)$ at the beginning of the two-year interval. They may be a subset of *AGE*, *HEIGHT*, *BAREROOT*, *CONTAINER*, *DAMAGE*, *DENSITY*, *AWDEN*, *SWDEN*, *PLDEN*, SC_{SP}, *SITE*, *DW*, and *HS*. Let y denote the binary response variable indicating a tree's life condition at the end of the two-year interval, where $y = 1$ if the tree was found alive, and $y = 0$ if the tree was found dead. There were n observations. n was equal to 7506 for white spruce, 8159 for lodgepole pine, and 8212 for aspen. The scatter plot of the response versus explanatory variables fell into two parallel lines without any hint of the relationship between mortality and the variables. In this situation, a special technique, called a generalized linear model (McCullagh and Nelder 1983), is needed.

Let $\pi(\boldsymbol{x})$ be the mathematical expectation of y under condition $\boldsymbol{x}$, E(y|$\boldsymbol{x}$), which is also the survival probability p in the two-year growth period, where $p = \Pr ob(y = 1 | \boldsymbol{x}) = \pi(\boldsymbol{x})$. The

idea of the generalized linear model is based on an assumption that a function of p may be expressed as a linear combination of $\boldsymbol{x}$. The formulation of the model is

$$g(p) = g\left(\pi(\boldsymbol{x})\right) = \beta_0 + x_1\beta_1 + x_2\beta_2 + \cdots + x_k\beta_k, \tag{3.1}$$

where $g(.)$ is a one-to-one differentiable function, called a link function, and β_0, β_1, $\cdots$, and β_k are unknown parameters (McCullagh and Nelder 1983). The link function relates the survival probability with the linear combination of the variables. Different link functions indicate different formulations of the model. The choice of a link function is usually made based on aptness to the data, model interpretation, and simplicity of the corresponding statistics (Hinkley *et al.* 1991). The most frequently used link function is the logistic link

$$g\left(\pi(\boldsymbol{x})\right) = \log\left(\frac{\pi(\boldsymbol{x})}{1-\pi(\boldsymbol{x})}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k, \tag{3.2}$$

and the corresponding generalized linear model is called the logistic model:

$$p = \frac{1}{1+\exp\left[-\left(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k\right)\right]}. \tag{3.3}$$

The logistic model is the cumulative distribution function of the logistic distribution, which is bounded by zero and one, the potential range of survival probabilities. With a selection of the proper set of variables and their transformations, the models should be able to describe most naturally occurring patterns of mortality (Hamilton 1990). The method for fitting the logistic model is called the logistic regression (Hosmer and Lemeshow 1989). It has been used for many tree species (Hann 1980, Buchman and Pederson 1983, Hamilton 1986, Hamilton 1990), since Hamilton (1974) and Monsured (1976) introduced it to model individual tree mortality,

Besides the logistic link, other two commonly used link functions are: the probit link: $g(p) = \Phi^{-1}(p)$, where Φ is the normal cumulative distribution function, and the compliment log-log link: $g(p) = \log\left(-\log(p)\right)$. The corresponding generalized linear models are the

probit model $p = \Phi(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)$, and the compliment log-log model $p = \exp[-\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)]$. The probit model is the cumulative normal distribution function, and the compliment log-log model is the cumulative extreme-value distribution function. A preliminary study showed that the probit model had a similar performance to the logistic model, and both of them were better than the compliment log-log model. We choose the logistic model in this study because of its broad application and acceptance.

3.3.2 Maximum likelihood estimation

Many statistical software packages provide specific procedures to fit the logistic model, such as the procedure PROC LOGISTIC and the procedure PROC PROBIT on SAS (SAS Institute Inc. 1992) and GLM on Splus (Statistical Science 1993). In this study the procedure PROC LOGISTIC on SAS (SAS Institute Inc. 1992) was used to fit the model with the maximum likelihood method for the parameter estimation, and Hosmer-Lemeshow test for the goodness-of-fit test.

The basic idea of the maximum likelihood estimation is as follows. The life condition code y is a binary response following a Bernoulli distribution when given a value of the input variable $\boldsymbol{x}$, thus for the i-th observation,

(3.4) $$\Pr ob(y_i = 1 | \boldsymbol{x}_i) = \pi(\boldsymbol{x}_i),$$

(3.5) $$\Pr ob(y_i = 0 | \boldsymbol{x}_i) = 1 - \pi(\boldsymbol{x}_i),$$

where y_i is the response variable for the i-th observation, $\boldsymbol{x}_i$ is the vector of the explanatory variable for the i-th observation. From (3.4) and (3.5), the probability distribution function of y_i can be written as:

(3.6) $$f(y_i, \pi(\boldsymbol{x}_i)) = \pi(\boldsymbol{x}_i)^{y_i} (1 - \pi(\boldsymbol{x}_i))^{1-y_i}.$$

For n observations, the likelihood function is calculated as

$$L(y,\pi(x)) = \prod_{i=1}^{n} \pi(x_i)^{y_i} (1-\pi(x_i))^{1-y_i}. \quad (3.7)$$

After a logarithm transformation of (3.7), the log likelihood function is

$$l(y,\pi(x)) = \log(L(y,\pi(x))) = \sum_{i=1}^{n} \{y_i \log(\pi(x_i)) + (1-y_i)\log(1-\pi(x_i))\}. \quad (3.8)$$

Based on the logistic link function $\log\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right) = \beta_0 + x_{i1}\beta_1 + \cdots + x_{ik}\beta_k$, the log likelihood function for the logistic model (3.3) is

$$l(y,\beta) = \sum_{i=1}^{n} \{y_i(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}) - \log(1+\exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}))\}. \quad (3.9)$$

To find the values of β that maximizes the log likelihood function (3.9), we differentiate (3.9) with respect to β, and set the resulting equations equal to zero. These equations are as follows:

$$\sum_{i=1}^{n} [y_i - (1+\exp(-(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik})))] = 0, \text{ and} \quad (3.10)$$

$$\sum_{i=1}^{n} x_{ij} [y_i - (1+\exp(-(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik})))] = 0, \text{ for } j = 1, 2, \cdots, k. \quad (3.11)$$

They are called the maximum likelihood equations, and are nonlinear in β. Numerical algorithms can be used to solve the nonlinear equation system, such as the Newton-Raphson method used in the procedure PROC PROBIT on SAS (SAS Institute Inc. 1992). For the logistic regression, it was shown that the solution of equations (3.10) and (3.11) can be obtained by a re-weighted least squares algorithm (McCullagh and Nelder 1983, page 31, Hinkley *et al.* 1991, page 62). In this approach, the weights are functions of the fitted survival probabilities. The process is iterative, because the weight depends on the fitted values, for which only current estimates are available.

This re-weighted least squares method is used in procedure PROC LOGISTIC, and the detailed computation steps can be found in SAS/STAT (SAS Institute Inc. 1992, page 1088).

3.3.3 Variable selection

The procedure PROC LOGISTIC (SAS Institute Inc. 1992) provide various variable selection methods, and one of them is a stepwise procedure. At one time, stepwise procedures were extremely popular in linear regression. However, in recent years there has been a shift away from deterministic methods for model building to purposeful selection of variables (Hosmer and Lemeshow 1989). In this study, both biologically meaningful selection and statistically significant selection were considered. In the mortality model, an appropriate biological interpretation of the mortality process is essential, because a model that is biologically illogical cannot necessarily be expected to perform well outside the data range (Hamilton 1986). In addition, the stepwise procedure still provides a useful tool when the relationship or significance of some variables are not known. In this study, information obtained from stepwise selection was combined with the understanding of the process of mortality to select an appropriate set of variables.

The stepwise procedure is based on a likelihood ratio test which checks for the importance of variables. Hosmer and Lemeshow (1989) explained the selection is that at any step in the procedure the most important variable will be one that produces the greatest change in the likelihood relative to a model not containing the variable. For example for the j-th variable with corresponding parameter β_j, the null hypothesis is that the variable can be eliminated, that is, H_0: $\beta_j = 0$. The likelihood ratio statistic is

$$(3.12) \qquad \Lambda = -2\log\left(\frac{L\left(y,\left(\tilde{\beta}_0,\tilde{\beta}_1,\ldots,\tilde{\beta}_{j-1},0,\tilde{\beta}_{j+1},\ldots\tilde{\beta}_k\right)'\right)}{L\left(y,\left(\hat{\beta}_0,\hat{\beta}_1,\ldots\hat{\beta}_{j-1},\hat{\beta}_j,\hat{\beta}_{j+1},\ldots,\hat{\beta}_k\right)'\right)}\right)$$

where $(\tilde{\beta}_0,\tilde{\beta}_1,\ldots,\tilde{\beta}_{j-1},0,\tilde{\beta}_{j+1},\ldots\tilde{\beta}_k)$ are the maximum likelihood estimates under H_0, and $\hat{\beta}=(\hat{\beta}_0,\hat{\beta}_1,\ldots,\hat{\beta}_k)$ are the maximum likelihood estimates under the full model. The likelihood ratio statistic Λ compares the likelihood without variable β_j and the likelihood with variable β_j. In a large sample, Λ is approximately distributed as a chi-square distribution with degree-of-freedom 1 under H_0. A large Λ value will provide evidence to against H_0.

3.3.4 Parameter significance test

Once the variables were selected and the model was fitted, the Wald chi-square test in the procedure PROC LOGISTIC (SAS Institute Inc. 1992) was used to assess the parameter significance. The Wald chi-square test was obtained by comparing the maximum likelihood estimate of the parameter, to an estimate of its standard error (Hosmer and Lemeshow 1989). The Wald chi-square statistic was computed as the square of the parameter estimate divided by its estimated standard error. Under the hypothesis that the parameter was not significantly different from zero, the Wald chi-square statistic should be a chi-square distribution with one degree-of-freedom.

3.3.5 Goodness-of-fit test

Once the model was fitted, the Hosmer-Lemeshow test, commonly employed in the logistic regression, was used to test the goodness-of-fit for the model. In the Hosmer-Lemeshow test, first, observations were sorted in an increasing order of their estimated survival probability, then the observations were divided into g groups. Ten groups was used in the procedure PROC

LOGISTIC (SAS Institute Inc. 1992). The Hosmer-Lemeshow goodness-of-fit statistic was obtained by calculating the Pearson chi-square statistic from the $2\times g$ table of observed and expected frequencies. A formula defining the calculation of the statistic was

$$x^2_{HW} = \sum_{i=1}^{g} \frac{\left(o_i - n_i \overline{p}_i\right)^2}{n_i \overline{p}_i \left(1 - \overline{p}_i\right)} \tag{3.13}$$

where n_i was the number of observations in the i-th group, o_i is the number of trees which were still alive at the end of two-year interval in the i-th group, and $\overline{p}_i$ was the averaged predicted survival probability for the i-th group. Under the assumption that no significant difference existed between the actual and predicted survival and mortality, the Hosmer-Lemeshow statistic should be approximately distributed as a chi-square distribution with degree-of-freedom $g-2=8$.

3.3.6 Modeling strategy

For each species, the mortality model (3.3) was fitted in the following steps:

1. Implement the stepwise selection procedure to the variables. Select the important variables based on 0.05 significant level for entry into or removal from the model.
2. Given the selected variables, estimate the unknown parameters using the maximum likelihood method through the re-weighted least squares algorithm.
3. Test the parameter significance based on the Wald chi-square test.
4. Check if the model provides the rational biological interpretations. If any relationship between a variable and the survival probability contradicts our understanding of the mortality process, go back to the original data, find the reason, consider deleting the variable or adding a transformation of the variable, and then go to step 2 with the new set of variables and transformations. Otherwise, go to next step.

5. Apply Hosmer-Lemeshow goodness-of-fit test to the model. If the model fails the goodness-of-fit test, list the chi-squares for the 10 groups (or Dbh groups, BA groups, ..., if needed), and find where the largest divergence is. Consider deleting or adding some variable transformations, and then go to step 2 with the new set of variables and transformations. Otherwise, go to next step.
6. Check again the biological interpretation of the model. Rational interpretation is much more important than the model precision.
7. Do steps 1 to 6 until all the results are satisfactory.

3.4 Results

Based on the two-year growth period, 7506 white spruce, 8212 aspen, and 8159 lodgepole pine were used for this analysis. The numerical variables *AGE*, *HEIGHT*, *DENSITY*, and SC_{SP} were summarized for both live and dead trees of each species (Table 3-1). The average age is 3 to 4 years. Average heights of spruce (16.7 cm) and pine (12.3 cm) are much less than the average height of aspen (102.5 cm). The average density is around 15,000 per ha. Species composition covers from 0 to 1.

Given the input values of the explanatory variables and the binary response, the procedure PROC LOGISTIC on SAS/STATS (SAS Institute Inc. 1992) was used to accomplish the fitting of the mortality model (3.3) (Appendix 3), following the 7 steps described in 3.3.6. The selected variables and transformations for white spruce, aspen, and lodgepole pine are:

White spruce: *AGE*, log(*HEIGHT*), $AWDEN^{-1}$, *BAREROOT*, *CONTAINER*, *SITE*, *HS* , *DAMAGE*;

Aspen: *AGE*, *HEIGHT*, $DENSITY^{-1}$, log(*DENSITY*), $SWDEN^{-1}$, log(*SWDEN*) and *SITE*;

Lodgepole pine: AGE, $HEIGHT$, $HEIGHT^2$, $\log(HEIGHT)$, $\log(DENSITY)$, $\log(PLDEN)$, SC_{PL}, $SITE$,

where SC_{PL} is lodgepole pine composition. The corresponding models are:

White spruce: (3.14)

$$P=\left[1+\exp\left(-\left(\beta_0+\beta_1 AGE+\beta_2\log(HEIGHT)+\frac{\beta_3}{AWDEN}+\beta_4 BAREROOT+\beta_5 CONTAINER+\beta_6 SITE+\beta_7 HS+\beta_8 DAMAGE\right)\right)\right]^{-1},$$

Aspen: (3.15)

$$P=\left[1+\exp\left(-\left(\beta_0+\beta_1 AGE+\beta_2 HEIGHT+\frac{\beta_3}{DENSITY}+\beta_4\log(DENSITY)+\frac{\beta_5}{SWDEN}+\beta_6\log(SWDEN)+\beta_7 SITE\right)\right)\right]^{-1},$$

and Lodgepole pine: (3.16)

$$P=\left[1+\exp\left(-\left(\beta_0+\beta_1 AGE+\beta_2 HEIGHT+\beta_3 HEIGHT^2+\beta_4\log(HEIGHT)+\beta_5\log(DENSITY)+\beta_6\log(PLDEN)+\beta_7 SC_{PL}+\beta_8 SITE\right)\right)\right]^{-1}.$$

The parameter estimates for models (3.14), (3.15), and (3.16) are listed in Tables 3-2, 3-4, and 3-6. All the parameters are significant based on the Wald chi-square test at level $\alpha=0.05$. The Hosmer and Lemeshow goodness-of-fit test for the three models are listed in Tables 3-3, 3-5, and 3-7 respectively. The values of the Hosmer and Lemeshow statistics for models (3.14), (3.15), and (3.16) are 6.995, 7.698, and 8.429. The corresponding p-values are 0.5372, 0.4635, and 0.3927 with 8 degree-of-freedom. They are large, far exceeding level $\alpha=0.05$ for all species, so there is no strong evidence showing the disagreement between the predicted and actual survival and mortality. The model is statistically well-fitted.

Since the data used to fit the models included multiple remeasurement intervals of the same individuals, the serial correlation might exist. To check the correlation, the data were sorted by time order for each individual. The empirical one-step serial correlation was examined by plotting each deviance residual (Hinkley *et al.* 1991, page 67) against the one preceding it for all individuals except the first measurement (Draper and Smith 1981). The one-step serial plot exhibited no significant tendency of the trend. To view the correlation for more steps, similar

plots for residuals two steps apart, three steps apart, and so forth, were made (Appendix 3). No significant tendency of the trend was found. In an empirical study of yield models, Borders *et al.* (1988) also found no serial correlation when the data derived from non-overlapping growth intervals.

3.5 Discussion

The models confirm some commonly held theories and relationships about white spruce, aspen, and lodgepole pine in Alberta regenerated mixedwood stands.

3.5.1 White spruce model

The negative coefficient of *DAMAGE* (-0.488572) suggests that damaged trees are susceptible to mortality. The positive coefficients of logarithm *HEIGHT* (1.310964) and *AGE* (0.397921) show that as a juvenile tree becomes older and taller, its ability to resist competition and damage agents is enhanced and the survival probability increases.

The coefficient of *BAREROOT* (-3.041297) is less than the coefficient of *CONTAINER* (-0.949001), which indicates that bareroot seedlings have lower survival probability than container seedlings when other variables are the same (Figure 3-1). This is compatible with the commonly accepted expectation that compared with the bareroot planting method, the container planting method is better for the root protection in seedling transportation and better for the root development after planting. For natural spruce seedlings, both variable *BAREROOT* and *CONTAINER* are equal to zero. Because of the negative coefficients of *BAREROOT* and *CONTAINER*, the natural spruce has higher survival probability in two-year period than the planted spruce when other variables are the same (Figure 3-1).

In aspen and spruce mixedwood stands, after clear-cutting, reproduction by root sucking assists the shade intolerant aspen to achieve fast early growth, high density, and rapid domination

of the site by occupying the upper layer of the canopy (Mueggler 1989). Aspen has a competitive advantage over the shade tolerant white spruce which exhibits slow juvenile growth. In the model of spruce, the positive coefficient of $AWDEN^{-1}$ (103.555128) shows that excessive aspen density may reduce white spruce survival. Besides aspen, grasses and shrubs may also play important roles in competing with white spruce seedlings for light, nutrients, and water. Due to the data limitation, those relationships were not analyzed here.

The positive coefficient of *SITE* (0.302884) and the negative coefficient of *HS* (-0.442657) suggest that white spruce survives better on richer sites. In this data set, the regenerated stands are younger than 15 years, the sites with better nutrient and water conditions may support better spruce early growth when the light supply is limited by the aspen canopy. The better growth may, then, assist the seedlings to resist damage agents and show higher survival probability, although more mortality on richer site may be expected later due to the better early growth which results in more competition among trees (Chapter 2).

3.5.2 Aspen model

In the aspen model, The positive coefficient of *HEIGHT* (0.0107) and *AGE* (0.3137) show that aspen seedlings have higher survival as they grow older and taller. The coefficient of $DENSITY^{-1}$ (5950.303916) and the coefficient of log(*DENSITY*) (0.0000610075) suggest that the density has a negative impact on aspen survival probability. Although aspen has the competitive advantage over shade tolerant white spruce, this shade intolerant species may still face competition among trees. Jones and Trujillo (1975) found that great densities commonly found in young aspen stands lead to either heavy juvenile mortality or severe growth stagnation. Normally, early differentiation into crown classes results in heavy mortality among overtopped aspen, and allows continued good growth by trees with strong canopy position (Jones and Trujillo 1975). As in the white spruce model, the positive coefficient of *SITE* (1.1405) shows that

aspen has high survival on richer site. In view of the relationship between aspen survival and spruce density, the negative coefficients of $SWDEN^1$ (-4408.206213) and log(SW*DEN*) (-1.772521) show that generally the presence of spruce also indicates higher survival of aspen, but an excessive spruce density may reduce aspen survival probability. More spruce may indicate richer sites, so aspen survives better as well. However with excessive spruce density, aspen may also face competition with spruce for nutrients and water and has a reduced survival probability, in spite of the fact that aspen has the competitive advantage for light.

3.5.3 Lodgepole pine model

All lodgepole pine trees used to build the model are from natural and seeded lodgepole pine stands. The positive coefficients of *AGE* (12.044120) and *HEIGHT* (0.318672) show that the taller and older trees have a competitive advantage over others and demonstrate higher survival probabilities. With increasing height, the negative coefficient of $HEIGHT^2$ (-0.001648) and the negative coefficient of log(*HEIGHT*) (-0.624499) slightly reduce the acceleration of survival probability, which is possibly due to more competition among the trees, but they will not change the trend that the survival probability increases as height increases. The negative impact of density is presented by the negative coefficient of log(*DENSITY*) (-2.3954). The negative coefficient of *SITE* (-0.7396) states that in natural and seeded lodgepole pine stands, faster growth on richer sites causes more competition among trees, and results in more likelihood of mortality. The positive coefficient of *PLDEN* (1.714658) and the negative coefficient of SC_{PL} (-5.06158) show that pine survives better in pure stands than in mixedwood stands.

In summary, the logistic model of mortality was fitted for white spruce, aspen, and lodgepole pine in Alberta mixedwood and pure juvenile stands. The model was statistically fitted well and logically presents biological process, and can be used to predict the survival probability

for individual trees by given age, height, stand density, species composition, and site productivity. In forest succession, mortality plays different roles. In the regeneration stage, mortality is relatively high. After establishment, stands show a fast growth and reduced mortality. When the growth space becomes limited, self-thinning begins, and the mortality increases. After self thinning, most survivors have a stable and high survival until old age. The data used in the study were from the regeneration stands younger than 15 years. The model only displays the relationship between mortality and total growth, competition, and site productivity for young stands. It should not be used to predict mature tree mortality.

Table 3-1. Summary statistics for juvenile tree and stand characteristics by life condition code

Species	Life condition	N	Variable	Mean	Std Dev	Minimum	Maximum
Spruce	Live	6979	*HEIGHT*	16.7525	17.4543	1.0000	169.0000
			AGE	3.8706	2.2637	1.0000	11.0000
			DENSITY	16021.05	1576.91	950.00	42500.00
			SC_{SW}	0.3707	0.2556	0.0044	0.9099
	Dead	527	*HEIGHT*	6.0152	7.9606	1.0000	43.0000
			AGE	1.9526	1.2515	1.0000	7.0000
			DENSITY	13416.35	6652.61	1200.00	20000.00
			SC_{SW}	0.3245	0.1911	0.0357	0.8380
Aspen	Live	7856	*HEIGHT*	102.5152	94.2395	2.0000	810.0000
			AGE	3.2351	1.9811	1.0000	10.0000
			DENSITY	16351.60	9257.15	950.00	42500.00
			SC_{AW}	0.2108	0.1526	0.0366	0.9677
	Dead	362	*HEIGHT*	50.3094	33.8941	1.0000	190.0000
			AGE	1.9006	0.7885	1.0000	5.0000
			DENSITY	14430.82	5580.32	1700.00	20350.00
			SC_{AW}	0.2475	0.1985	0.0516	0.7121
Pine	Live	7410	*HEIGHT*	12.3120	16.7487	1.0000	186.0000
			AGE	3.2049	2.0275	1.0000	9.0000
			DENSITY	14944.55	6758.22	250.00	27850.00
			SC_{PL}	0.8665	0.1333	0.0109	1.0000
	Dead	749	*HEIGHT*	2.8278	2.1928	1.0000	21.0000
			AGE	1.4045	0.7743	1.0000	5.0000
			DENSITY	16425.52	6022.94	1881.00	22800.00
			SC_{PL}	0.9248	0.1032	0.0426	1.0000

Table 3-2. Parameter estimates for white spruce mortality model (3.14)

Variable	Parameter	Parameter estimate	Standard error	Wald Chi-square	Pr > Chi-square
INTERCEPT	β_0	0.166788	0.1014	2.7074	0.0490
AGE	β_1	0.397921	0.0623	40.8103	0.0001
log(*HEIGHT*)	β_2	1.330964	0.1095	143.2225	0.0001
$AWDEN^{-1}$	β_3	103.555128	32.7005	10.0285	0.0015
BAREROOT	β_4	-3.041297	0.2435	155.9826	0.0001
CONTAINER	β_5	-0.949001	0.2535	14.0155	0.0002
SITE	β_6	0.302884	0.1259	5.7909	0.0161
HS	β_7	-0.442657	0.2127	4.3304	0.0374
DAMAGE	β_8	-0.488573	0.1433	11.6238	0.0007

Table 3-3. Hosmer and Lemeshow goodness-of-fit test for white spruce model (3.14)

		Number of live trees		Number of dead trees	
Group	Total number of trees	Observed	Expected	Observed	Expected
1	751	502	503.98	249	247.02
2	751	656	643.17	95	107.83
3	751	680	687.17	71	63.35
4	751	707	708.91	44	42.09
5	751	717	723.41	34	27.59
6	751	735	732.65	16	18.35
7	751	739	739.89	12	11.11
8	751	746	744.97	5	6.03
9	751	750	748.22	1	2.78
10	751	747	746.16	0	0.84

Hosmer-Lemeshow statistic = 6.995 with 8 DF (p=0.5372)

Table 3-4. Parameter estimates for aspen mortality model (3.15)

Variable	Parameter	Parameter estimate	Standard error	Wald Chi-square	Pr > Chi-square
INTERCEPT	β_0	16.159941	1.7031	87.9562	0.0001
AGE	β_1	0.313703	0.0631	24.7064	0.0001
HEIGHT	β_2	0.010747	0.0018	34.5411	0.0001
DENSITY^{-1}	β_3	5950.303916	1265.3121	22.1161	0.0001
SWDEN^{-1}	β_4	-4408.206213	450.3002	95.8062	0.0001
log(*DENSITY*)	β_5	0.000061	0.000015	16.9244	0.0001
log(*SWDEN*)	β_6	-1.772521	0.2066	73.6018	0.0001
SITE	β_7	1.140481	0.2288	24.8495	0.0001

Table 3-5. Hosmer and Lemeshow goodness-of-fit test for aspen model (3.15)

		Number of live trees		Number of dead trees	
Group	Total number of trees	Observed	Expected	Observed	Expected
1	822	698	703.50	124	118.50
2	822	746	749.08	76	72.92
3	822	761	766.63	61	55.37
4	822	784	780.98	38	41.02
5	822	801	791.96	21	30.04
6	822	805	801.23	17	20.77
7	822	809	809.56	13	12.44
8	822	812	815.03	10	6.97
9	822	821	818.83	1	3.17
10	820	819	819.18	1	0.82

Hosmer-Lemeshow statistic = 7.6981 with 8 DF (p=0.4635)

Table 3-6. Parameter estimates for lodgepole pine mortality model (3.16)

Variable	Parameter	Parameter estimate	Standard error	Wald Chi-square	Pr > Chi-square
INTERCEPT	β_0	12.044120	1.1464	110.3683	0.0001
AGE	β_1	0.438379	0.0744	34.7584	0.0001
HEIGHT	β_2	0.318672	0.0563	31.9887	0.0001
$HEIGHT^2$	β_3	-0.001648	0.0004	18.4454	0.0001
log(*HEIGHT*)	β_4	-0.624499	0.1635	14.5812	0.0001
log(*DENSITY*)	β_5	-2.395358	0.3943	36.9094	0.0001
log(*PLDEN*)	β_6	1.714658	0.3843	20.0023	0.0001
SC	β_7	-5.061580	0.9042	31.3340	0.0001
SITE	β_8	-0.739634	0.1816	16.5879	0.0001

Table 3-7. Hosmer and Lemeshow goodness-of-fit test for lodgepole pine model (3.16)

		Number of live trees		Number of dead trees	
Group	Total number of trees	Observed	Expected	Observed	Expected
1	816	574	589.18	242	226.82
2	816	635	634.90	181	181.10
3	816	699	677.24	117	138.76
4	816	729	729.38	87	86.62
5	816	752	756.35	64	59.65
6	816	783	780.47	33	35.53
7	816	795	800.21	21	15.79
8	816	812	811.82	4	4.18
9	816	816	815.46	0	0.54
10	815	815	814.92	0	0.08

Hosmer-Lemeshow statistic = 8.4291 with 8 DF (p=0.3927)

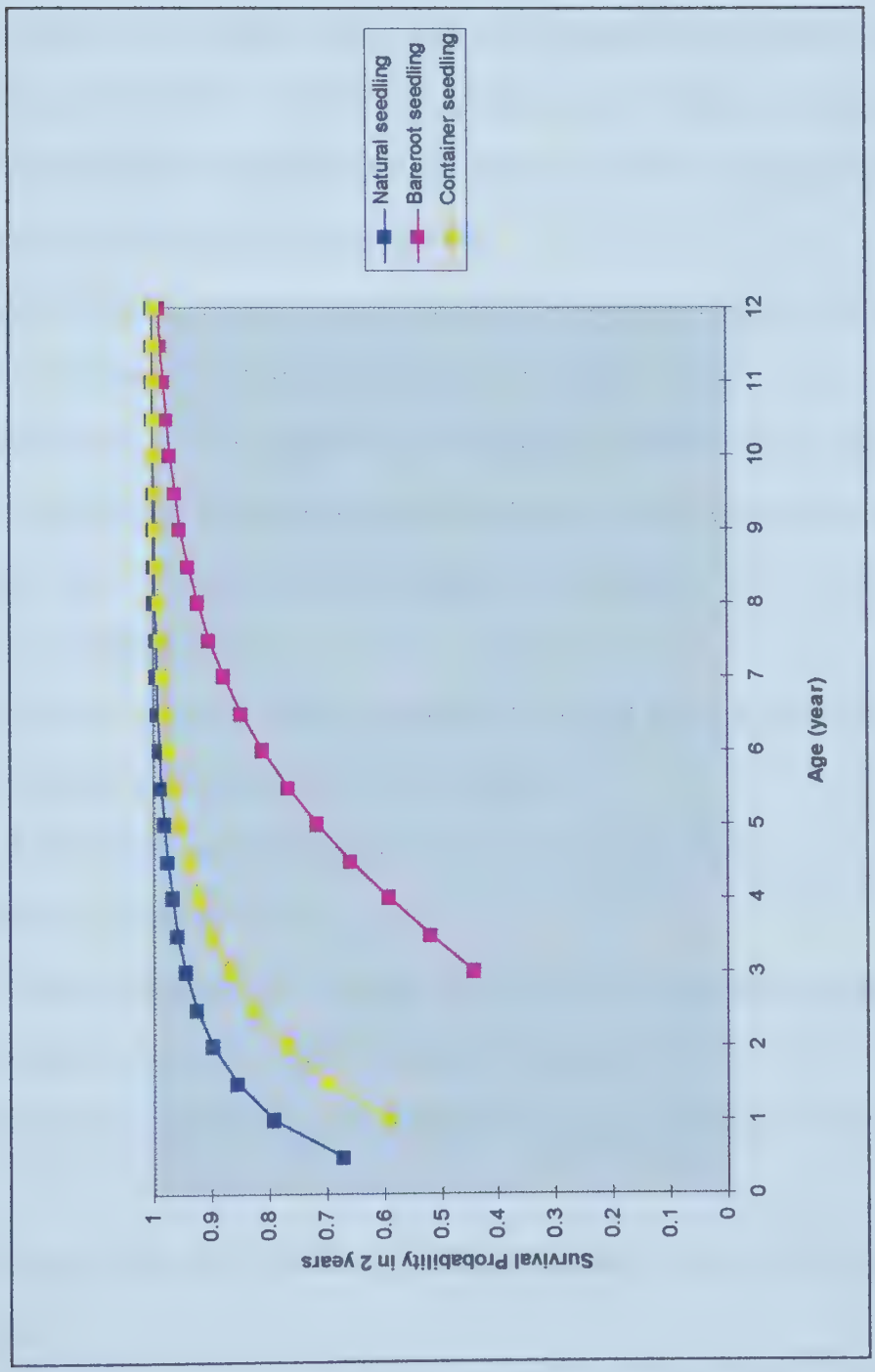


Figure 3-1. The predicted survival probability in two years for white spruce seedlings, given medium site, 10000/ha aspen, and 5000/ha white spruce

3.5 References

Barbour, M.G., Pavlik, B.M., and Antos, J.A. 1990. Seedling growth and survival of red and white fir in a Sierra Nevada ecotone. Amer. J. Bot. 77(7):927-938.

Border, B.E., Bailey, R.L. and Clutter, M.L., 1988. Forest growth models: Parameter estimation using real growth series. In A.R. Ek, S.R. Shifley and T.E. Burk (eds) forest Growth Modeling and Prediction. Proc. IUFRO Conf., 23-27 Aug. 1987, Minneapolis, MN. USDA For. Serv., Gen. Tech. Rep. NC-120, pp. 660-667.

Buchman, R.G., Pederson, S.P., and Walters, N.R. 1983. A tree survival model with application to species of the Great Lakes region. Can. J. For. Res., 13:601-608.

Draper, N.R. and Smith, H. 1981. Applied regression analysis. John Wiley & Sons, Inc.

Gilday, D.W. 1990. Monitor Plot survey manual. Published by Alberta Land and Forest Service.

Hamilton, D.A. 1974. Event probabilities estimated by regression. USDA Forest Service research paper. INT-152, 18p.

Hamilton, D.A. 1986. A logistic model of mortality in thinned and unthinned mixed conifer stands of Northern Idaho. Forest Science, 32:989-1000.

Hamilton, D.A. 1990. Extending the range of applicability of an individual tree mortality model. Can. J. For. Res., 20:1212-1218.

Hann, D.W. 1980. Development and evaluation of an even- and uneven-aged ponderosa pine / Arizona fescue stand simulator. U.S.D.A. For. Serv. Res. Pap. INT-267.

Hinkley, D.V., Reid, N., and Snell, J. 1991. Statistical theory and modeling. Chapman and Hall Ltd.

Hosmer, D.W. and Lemeshow, S. 1989. Applied logistic regression. John Wiley & Sons. New York. USA.

Jones, J. R. and Trujillo, D.P. 1975. Development of some young aspen stands in Arizona. USDA rocky mountain forest and range experiment station. Research paper RM 151. 11p.

Maddala, G.S. 1988. Introduction to econometrics. Macmillan Publishing company, New York. Collier Macmillan Publisher, London.

McCullagh, P., and Nelder, J.A. 1983. Generalized linear models. Chapman and Hall Ltd.

Monserud, R.A. 1976. Simulation of forest tree mortality. Forest Science, 22:438-444.

Morse, B.W. and Kulman, H.K. 1984. Plantation white spruce mortality: estimates based on aerial photography and analysis using a life-table format. Can. J. For. Res. 14:195-200.

Mueggler, W.F. 1989. Age distribution and reproduction of intermountain aspen stands. Western J. Appl. For.: 4(2):41-46.

SAS Institute Inc. 1992. SAS user's guide: statistics. Version 6 ed., Cary, NC 27512-8000.

SAS Institute Inc. 1992. SAS technical report P-229, SAS/STAT software: Changes and enhancement. Cary, NC 27513.

Shiver, B.D., Borders, B.E., Page, H.H., Jr., and Raper, S.M. 1990. Effect of some seedling morphology and planting quality variables on seedling survival in the Georgia Piedmont. South. J. Appl. For. 14(3):109-114.

Statistical Science. 1993. S-PLUS Guide to statistical and mathematical analysis, version 3.2. Seattle: StatSci, a division of mathsoft Inc.

Turner, I.M. 1990. Tree seedling growth and survival in a Malaysia rain forest. Biotropica. 22(2):146-154.

Vanclay. J.K. 1994. Modeling forest growth and yield - Applications to mixed tropical forests. CAB international, Wallingford, Oxon OX10 8DE, UK.

Wakeley, P.C. 1954. Planting the southern pines. USDA Monogr. 18. 233p.

Chapter 4

A lifetime distribution model

4.1 Introduction

During the last several decades, extensive mixedwood stands in Alberta have been harvested and replaced by juvenile stands. As these regenerated stands become more and more important for forest management in the provincial area, the study of juvenile tree mortality is imperative in terms of understanding mortality process, guiding silvicultural activities, and contributing to stand growth prediction.

In studies of survival probability for individual trees, logistic regression is a commonly used technique. It has been applied to many tree species in both pure and mixedwood stands (Hamilton 1974, Monserud 1976, Hann 1980, Buchman and Pederson 1983, Hamilton 1986, and Hamilton 1990). In a fitted logistic mortality model, the survival probability in a given time period can be predicted based on characteristics of the tree and stand. However, to make a longer term prediction, the model has to be iterated with updated growth predictions in each growth period.

Lifetime distribution analysis (Lawless 1982, Kalbfleisch 1980, Lee 1992) can portray the whole lifetime distribution instead of the survival in a given time period. The lifetime, or survival time, defined as the elapsed time from the initial measurement to death, is a random variable. It may be subject to many biotic and abiotic, internal and external factors. Once an appropriate statistical distribution model has been constructed for the lifetime and its parameters estimated, the survival probability can be predicted in any time period.

A parametric regression model relates the parameters of the lifetime distribution to explanatory variables. Since the lifetime analysis focuses on the lifetime distribution and the

logistic regression only gives emphasis to the periodic survival probability, the variables which are important in one model may be insignificant in the other. For example, some damage agents do not kill a tree in a few years, but shorten the survival time of the tree. Therefore the lifetime distribution analysis may be more sensitive to some causes of mortality.

Lifetime distribution analysis has developed into an important topic in biomedical and reliability research (Nelson 1990, Lee 1992). In forest growth and yield, the exponential distribution (Moser 1972) and Weibull distribution (Somers *et al.* 1980) were used to fit the survival as stand level models. For individual tree survival, only Maguire and Maguire (1987) implemented a parametric extreme value regression model to Douglas-fir seedling survival. Since the data only included the first year survival, the interpretation and prediction were limited.

The parametric regression method requires an understanding of the lifetime distribution. Unfortunately most of the lifetime distributions are not simple, and the data requirement is substantial. In the logistic regression, the data is obtained as repeated measurements on many individuals. In a lifetime distribution analysis, individuals have to be traced for a relatively long time, if possible, until death. Theoretically the parametric regression model is more difficult to fit, and practically less direct to apply than the logistic regression. This may explain why logistic regression is more common and few parametric regression models of lifetime distribution have been found in tree mortality studies.

In this paper, an application of lifetime distribution analysis is illustrated. The data are permanent sample plots established on cut-blocks after harvesting. All the individual trees have been traced and measured up to 15 years. The Weibull distribution and log-normal distribution were used to fit the lifetime distribution for Alberta major tree species: white spruce (*Picea glauca* (Moench) Voss), lodgepole pine (*Pinus contorta* var. *latifolia* Engelm), and aspen (*Populus tremuloides* Michx). A parametric regression model was constructed, which relates the lifetime distribution of individuals to tree age, height, damage code, regeneration origin, and

stand density and site productivity. The relationships between survival and impacting factors are discussed, and the examples of the model usage are shown.

4.2 Materials

4.2.1 Data

The Alberta Land and Forest Services provided the data for this study. The data are permanent sample plots designed to "monitor" juvenile stand establishment and growth on cut-blocks (Gilday 1990). The plots have been established in various ecoregions across the province since 1980. There are 107 plots; each plot has been remeasured every two years and up to 7 times. At plot establishment, harvest, scarification, regeneration treatment, ecoregion, moisture regime, drainage condition, duff depth and other stand characteristics were recorded for each plot. The regeneration treatments included left-for-natural, seeded, and planted with different species. Of the 107 plots, 69 were planted with white spruce; 22 were left for natural or seeded with lodgepole pine or mixed species; 7 were planted with lodgepole pine; 3 were seeded with white spruce; and 6 were seeded with forage and unspecified species. The last three groups were not of sufficient size for analysis and 6 plots have been re-seeded or thinned. They were excluded from this study.

Each plot consisted of 40 2 m × 2.5 m quadrats that were alternatively established on each side of four 25 metre lines distributed over a cut-blocks (Gilday 1990). Lodgepole pine, jack pine, white spruce, black spruce, and all fir species were defined as acceptable species. In a quadrat, each tree of acceptable species and the three tallest deciduous were tagged and measured with respect to species, age, height, origin, damage code, and etc. Other deciduous trees were counted by species.

4.2.2 Lifetime and explanatory variables

Each tree was a single observation for the analysis. When a tree was first tagged in the field, the tree and stand characteristics at the first measurement time was recorded as the value of explanatory variables. Later, if the tree was found dead in remeaurements, the death year was identified in field and the lifetime, the elapsed time from the first measurement to death, was computed. Trees dead before the last measurement were called completed observations and their lifetimes were known. Trees still alive at the time of the last measurement were called censored observations and the elapsed time to the last measurement was a censored life time. For each observation, an indicator variable *CENSOR* was introduced, which was equal to 0 if the tree was a completed observation, or 1 if it was a censored observation.

In the data, the damage code was used to indicate the damage categories for live trees and to report possible causes of mortality when trees were first found dead on the site (Gilday 1990). It included 21 categories such as suppression, flooding, climatic extremes, diseases, insects, and etc., which are listed in the first two columns of Table 4-1. In this study, the damage code was considered as an explanatory variable (*DAMAGE*).

Tree regeneration origin code (*ORIGIN*) had three classes: bareroot for trees which were nursery grown bareroot seedlings, container for trees which were nursery grown containerized seedlings, and ingress for trees which were germinated from natural or applied seeding on the site. Since all cut-blocks were harvested by clear-cutting, a small number of advance trees, those which had grown on the site for at least one year prior to harvesting, were excluded from this study.

Tree age and height were very important to indicate the tree's growth, vigor, and the ability to resist internal and external damage agents. In the data set, both *AGE* and *HEIGHT* (cm) were measured, and considered as explanatory variables.

Although not all the deciduous were tagged and remeasured, all of them were counted by species. This made it possible to summarize total density, density by species, and species

composition as additional variables. Total density (*DENSITY*) was defined as the total number of trees for all species per ha. *AWDEN*, *SWDEN*, and *PLDEN* were aspen density, spruce density, and lodgepole pine density. Species composition (SC_{SP}) was equal to the number of trees per ha for a target species divided by number of trees per ha for all species.

The site productivity classification of cut-blocks was identified from the Phase 3 Inventory. Since almost all the cut-blocks were located in good and medium sites, the site class only had two categories: good and medium. In addition, the drainage condition, an assessment of soil drainage characteristics of each block, and the humus depth, the average depth of the duff layer on the block, were also considered. In the original data, the drainage condition had 12 classes and humus depth had 5 classes. In this study the drainage classes (*DW*) were merged into well or poor, two groups. The humus depth (*HS*) was divided into deep ($\geq$ 16 cm) and shallow ($<$ 16 cm).

Among the data, 3653 white spruce, 4136 aspen, and 4056 lodgepole pine seedlings were used to the analysis. The summary statistics for each species are listed in Table 4-2.

4.3 Methods

Let T be the lifetime of an individual tree. T is a nonnegative random variable and is subject to many internal and external agents. The main objective here is to fit a lifetime distribution model to the survival data. The distribution of lifetime T is usually described or characterized by three functions: the probability distribution function, the survival function, and the hazard function. The three functions are mathematically equivalent. If one is known, the other two can be derived. An essential issue in the lifetime distribution analysis is to estimate one or more of these functions and to draw inferences about the survival pattern (Lee 1992).

1. Probability distribution function. Like any other continuous random variable, T has a probability distribution function defined as

$$f(t) = \lim_{\Delta t \to 0} \frac{P(\text{an individual tree dying in the interval } (t, t+\Delta t))}{\Delta t} \quad (4.1)$$
$$= \lim_{\Delta t \to 0} \frac{P(t \le T \le t+\Delta t)}{\Delta t}$$

It has the properties: $f(t) \ge 0$, and $\int_0^{+\infty} f(t)dt = 1$.

2. Survival function. The survival function is defined as the probability of an individual tree surviving longer than *t*:

$$S(t) = P(\text{an individual tree surviving longer than t}) \quad (4.2)$$
$$= P(T > t) = \int_t^{\infty} f(x)dx.$$

$S(t)$ is a monotone decreasing function with $S(0)=1$ and $S(\infty)=0$, that is, the probability of surviving at time zero is 1 and that of surviving an infinite time is zero. The annual survival probability (P_S) of the individual, commonly used in the logistic regression, can be calculated based on the survival function:

$$P_s(t) = P(T \ge t+1 | T \ge t) = \frac{S(t+1)}{S(t)}. \quad (4.3)$$

Since $S(t)$ is monotone decreasing and nonnegative, $P_s(t)$ is bound by 0 and 1 as expected.

3. *Hazard function*. The hazard function is defined as the instantaneous rate of mortality at time *t*, given that the individual tree survives until t, that is

$$h(t) = \lim_{\Delta t \to 0} \frac{P\{\text{an individual of time } t \text{ dying in the time interval } (t, t+\Delta t)\}}{\Delta t} \quad (4.4)$$
$$= \lim_{\Delta t \to 0} \frac{P(t \le T \le t+\Delta t | T \ge t)}{\Delta t} = \frac{f(t)}{S(t)}.$$

The hazard function denotes the conditional probability per unit time of dying. It describes the mortality risk as time passes. For different lifetime distribution, the shapes of the hazard functions are qualitatively quite different. The hazard function may increase, decrease, remain

constant, or indicate a more complicated process. Individual trees in stands are followed right from birth to death. A so-called bathtub or U-shaped hazard function is often appropriate to describe the death risk of individual trees. It presents the high risk of death in young and old ages and low risk of mortality in middle ages. However in this study, all the plots were located in the regenerated cut-blocks and most trees were younger than 15 years. Therefore the U-shaped hazard function is not suitable.

Among the lifetime distribution models, exponential, Weibull, gamma, and log-normal distributions occupy a central role because of their applications in a wide range of situations. We have little knowledge about the juvenile tree lifetime distribution. The basic question is: what lifetime distribution model should be used? An inappropriate distribution model will result in a biased survival prediction. Lee (1992) suggested using nonparametric methods to analyze survival data before attempting to fit a theoretical distribution. The estimates and graphs obtained by nonparametric methods may help in choosing a distribution model.

4.3.1 Nonparametric estimation

Nonparametric or distribution-free methods are a useful way to portray the survival data by computing and graphing the empirical survival function, empirical probability distribution function, and empirical hazard function. The followings describe the nonparameteric methods used in this study.

Suppose trees are observed dead at times $t_1 < t_2 < \cdots < t_k$. More than one dead trees at the same time is allowed. There are also censored trees whose lifetimes are not observed because they don't die before their last measurements. At time t_i, the empirical survival function is estimated by the product-limit estimation (Lawless 1982, P71), which is defined as

$$\hat{S}(t_i) = \prod_{j=1}^{i} \frac{n_j - d_j}{n_j} \qquad (4.5)$$

where n_j is the number of individuals alive and uncensored just prior to t_j and d_j is the number of deaths at t_j. The estimate $\hat{S}(t_i)$ is built up as a product, and each term of the product can be thought of as an estimate of the conditional probability of survival past time t_j, given survival until just prior to t_j. The product-limit estimates are essentially equivalent to the life-table estimates which is one of the oldest methods in tree mortality study (Harcombe 1987).

The probability distribution function can be estimated (SAS Institute Inc. 1992) as

$$\hat{f}(t_i) = \frac{P(t_i \le T < t_{i+1})}{t_{i+1} - t_i} = \frac{P(T \ge t_i) P(t_i \le T < t_{i+1} | T \ge t_i)}{t_{i+1} - t_i} = \frac{\hat{S}(t_{i-1}) \frac{d_i}{n_i}}{t_{i+1} - t_i}. \quad (4.6)$$

The numerator is the estimated probability of a tree dying in a time interval and the denominator is the magnitude of the interval. The hazard function can be estimated as (Lawless 1982, p8)

$$\hat{h}(t_i) = \frac{\hat{f}(t_i)}{\hat{S}(t_i)}. \quad (4.7)$$

The above method is quite easy to understand and apply. It is more efficient than parametric methods when no suitable theoretical distributions are known and less efficient when survival time follows a theoretical distribution (Lee 1992). It provides the nonparametric estimates of the three functions. Graphing the empirical functions, the survival pattern can be observed.

4.3.2 Theoretical distribution models

Exponential, Weibull, gamma, and log-normal distributions are commonly used models to fit the lifetime data. In this study, only Weibull distribution and log-normal distribution are considered.

The Weibull distribution model is fairly flexible, and it has simple expressions for the probability distribution function, survival function, and hazard function. The hazard function can

be decreasing, increasing, or constant. Assume that the lifetime T follows a Weibull distribution with a scale parameter α and shape parameter δ. The probability distribution function, survival function, and hazard function were summarized in Lawless (1982) and they are listed in Table 4-3.

The log-normal distribution is most easily specified by saying that the lifetime T is log-normally distributed if the logarithm of the lifetime, $Y = \log T$, is normally distributed. The hazard function of T increases initially to a maximum and then decreases to zero as time approaches infinity. Suppose that T is a log-normal distribution and Y has mean μ and standard deviation σ. The probability distribution function, survival function, and hazard function of T are listed in Table 4-3, where $\Phi(.)$ is the cumulative distribution function for a standard normal distribution.

The exponential distribution has a constant hazard function and is often referred to as a pure random failure pattern (Lee 1992). The gamma distribution describes a survival pattern where the hazard function is decreasing or increasing to a constant value as time passes. These two distributions are not considered in this study because the hazard rate of juvenile trees is not expected to be a constant or to approach a constant.

4.3.3 Parametric regression models

When the Weibull distribution or log-normal distribution is chosen to describe the survival pattern, we need to estimate the parameters of the distributions from the data. The parametric regression models allow us to relate the explanatory variables to the parameters of the lifetime distribution. The main idea is as follows. Under a theoretical lifetime distribution, the log lifetime distribution can be derived. Let $\boldsymbol{x} = (x_1, x_2, \cdots, x_k)$ be explanatory variables, and μ and σ denote the location parameter and the scale parameter of the log lifetime $(\log T)$

distribution. We assume only that μ is a function of explanatory variables $\boldsymbol{x}$, and that σ is a constant. In this study the linear form of $\boldsymbol{x}$ was chosen, which is

$$\mu(\boldsymbol{x}) = \boldsymbol{x}\beta \tag{4.8}$$

where $\beta = (\beta_1, \beta_2, \cdots, \beta_k)'$ is a $k\times 1$ vector unknown parameters. By estimating the parameters β and σ based on the maximum likelihood method, the parameters of the log lifetime distribution and the lifetime distribution can be specified.

If T is a Weibull distribution with a scale parameter α and shape parameter δ, then $Y = \log(T)$ is the extreme value distribution with a location parameter $\mu(\boldsymbol{x})$ and a constant scale parameter σ (Lawless 1982), where $\delta = 1/\sigma$ and $\alpha = \exp(\mu(\boldsymbol{x}))$. Given $\boldsymbol{x}$, the probability distribution function and the survival function of Y are listed in Table 4-4. The constancy of σ corresponds to the constancy of δ. Only α depends on $\boldsymbol{x}$.

If T is a log-normal distribution, then Y is the normal distribution with a location parameter $\mu(\boldsymbol{x})$ and a constant scale parameter σ. Given $\boldsymbol{x}$, the probability distribution function and the survival function of Y are listed in Table 4-4.

4.3.4 Maximum likelihood method

To estimate the parameters β and σ, the maximum likelihood method was used in this study. In the data, through the first measurement to the last measurement, some seedlings died which had the complete lifetime, but some of them were still alive at the last measurement. Let D denote the set of individuals with complete lifetime, and C be the set of individuals with censored lifetime. Observations $(y_i, \boldsymbol{x}_i)$, $i=1,2,...,n$, are available, where y_i is either a log lifetime or a log censored time, and $\boldsymbol{x}_i$ is the vector value of explanatory variables. The likelihood function is

$$L(\beta,\sigma)=\prod_{i\in C} f(y_i|\boldsymbol{x}_i)\prod_{i\in D} S(y_i|\boldsymbol{x}_i). \tag{4.9}$$

Let r be the number of trees having complete lifetime. For the Weibull lifetime distribution, based on the probability distribution function and survival function of $\log(T)$ provided in Table 4-4, the log likelihood function is (Lawless 1982, p299)

$$\log L(\beta,\sigma)=-r\log\sigma+\sum_{i\in D}\frac{y_i-\boldsymbol{x}_i\beta}{\sigma}-\sum_{i=1}^{n}\exp\left(\frac{y_i-\boldsymbol{x}_i\beta}{\sigma}\right). \tag{4.10}$$

For the log-normal lifetime distribution, based on the probability distribution function and survival function of $\log(T)$ provided in Table 4-4, the log likelihood function is (Lawless 1982, p314)

$$\log L(\beta,\sigma)=-r\log\sigma-\frac{1}{2\sigma^2}\sum_{i\in D}(y_i-\boldsymbol{x}_i\beta_i)^2+\sum_{i\in C}\log\left(1-\Phi\left(\frac{y_i-\boldsymbol{x}_i\beta_i}{\sigma}\right)\right). \tag{4.11}$$

The log likelihood functions contain the unknown parameter β and σ. To maximize the log likelihood functions, that is, to solve the maximum likelihood equations: $\frac{\partial \log L}{\partial\beta}=0$ and $\frac{\partial \log L}{\partial\sigma}=0$, the estimates $\hat{\beta}$ and $\hat{\sigma}$ can be obtained.

4.3.5 Parameter significance test

Once the model is fitted, the parameter significance is to be assessed. The likelihood ratio method was used (Lawless 1982) to test the hypothesis H_0: $\beta_j=0$ versus H_1: $\beta_j\neq 0$. The likelihood ratio statistic is

$$\Lambda=-2\log\left(\frac{L\left(y,\left(\tilde{\beta}_0,\tilde{\beta}_1,\ldots,\tilde{\beta}_{j-1},0,\tilde{\beta}_{j+1},\ldots\tilde{\beta}_k;\tilde{\sigma}\right)'\right)}{L\left(y,\left(\hat{\beta}_0,\hat{\beta}_1,\ldots\hat{\beta}_{j-1},\hat{\beta}_j,\hat{\beta}_{j+1},\ldots,\hat{\beta}_k;\hat{\sigma}\right)'\right)}\right) \tag{4.12}$$

where $\tilde{\beta}_0, \tilde{\beta}_1, \cdots, \tilde{\beta}_k$ and $\tilde{\sigma}$ are the maximum likelihood estimates under H_0: $\beta_j=0$, and $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_k)$ and $\hat{\sigma}$ are the maximum likelihood estimates of σ and β under the full model. Assume $\hat{\beta}_j$ is an l dimensional vector. In a large sample, Λ is approximately distributed as an $\chi^2_{(l)}$ distribution under H_0. A large Λ value provides evidence to against H_0 and shows that β_j is significant.

4.3.6 Residual analysis

After the parametric regression model is fitted under a theoretical distribution assumption, the analysis of residual ensures that the assumption is acceptable and the model is not biased.

For the Weibull distribution, the residual is defined as

$$e_i = \left(t_i \exp\left(-\boldsymbol{x}_i \beta\right)\right)^{\delta}. \tag{4.13}$$

The residual e_i follows a standard exponential distribution (Lawless 1982, p280) under the assumption that lifetime T is a Weibull distribution. Plotting the ordered residuals $e_{[i]}$ against α_i, where

$$\alpha_i = \sum_{j=1}^{i} (n - j + 1)^{-1}, \tag{4.14}$$

the plot should be approximately linear with slope 1.

For log-normal distribution model, the residual is defined as

$$e_i = \frac{\log(t_i) - \boldsymbol{x}_i \beta}{\sigma}, \tag{4.15}$$

which should be a standard normal distribution if the assumption that lifetime T is a log normal distribution is suitable.

4.4 Results and Discussion

4.4.1 Possible causes for mortality

When trees were first found dead on the site, the damage code reported possible causes of the mortality. Therefore the mortality trees were summarized by damage code for each species and listed in Table 4-1. The possible causes of mortality, and the number and percentage of dead trees caused by each damage agent were listed. It was found that suppression was a common cause of juvenile tree mortality, where the percentages of dead trees caused by suppression for spruce, lodgepole pine, and aspen were 22.59%, 18.21%, and 23.23%. Climatic extremes caused mortality for spruce (11.85%) and pine (17.33%), but was less important for aspen (3.73%). Of all the mortality, 19.10% spruce mortality was brought about by flooding. Herbicide (35.06%), and diseases (11.61%) were harmful to aspen seedling survival. The causes of nearly half (49.12%) of the pine mortality were unknown.

The damage code (*DAMAGE*) was an explanatory variable. Based on the summary of possible mortality causes, the damage code was initially grouped into 6 classes: suppression, climatic extremes, flooding, herbicide, other damage, and no damage.

4.4.2 Nonparametric estimation

Based on equation (4.5), (4.6), and (4.7), the procedure PROC LIFETEST on SAS (SAS Institute Inc. 1992) was used to calculate the nonparametric estimates of the survival function, probability distribution function, and hazard function for each species (Appendix 4). The estimated hazard functions against time for each species are shown in Figure 4-1. It was observed that for natural lodgepole pine and aspen seedlings, after regeneration, the hazard rate increased, reached the maximum value around 2 to 3 years, and then decreased. If a seedling survived the first 6 years, it experienced a very low mortality risk in later years. The log-normal distribution has a unimodal shape hazard function that initially increases to a maximum and then decreases to

zero as time passes. It was chosen to fit aspen and pine lifetime distributions. In Figure 4-1, for either bareroot or container planted spruce, the hazard rate reaches the maximum value immediately after planting, and then monotonously declines with the passage of time. When the scale parameter $\delta < 1$, the Weibull distribution has a decreasing hazard rate that is compatible with the planted spruce. For natural spruce, the hazard trend is similar to aspen and pine because of the natural origin. Since all the spruce are in plantations, and the number of spruce trees would be too few if the planted and natural trees were analyzed separately, both Weibull and log-normal distribution were considered as candidates for spruce lifetime distribution model. If the Weibull distribution is chosen, it is expected that the scale parameter $\delta < 1$.

4.4.3 Parametric regression model

The procedure PROC LIFEREG on SAS (SAS Institute Inc. 1992) was used to fit the parametric regression model based on the maximum likelihood method (Appendix 4). The selection of explanatory variables were made based on biological appropriateness and parameter significance tests. The likelihood ratio method was used to test the significance of the parameters. It was found that the impact from the number of trees by species was more significant than that of the species composition. Initially *DAMAGE* had 6 classes. The preliminary fitting showed that the parameters of the 6 classes were not always significant, that is, the mortality pattern by different damage agents might not be significantly different. Therefore the damage classes were merged according to the significance test. They are described below as part of the estimation discussion.

In the spruce lifetime distribution model, it was assumed that the lifetime T follows a Weibull distribution. Then the extreme value distribution was fitted to the log lifetime Y. The selected variables for equation (4.8) were *HEIGHT*, *AWDEN*^{1}, *DENSITY*^{1}, *ORIGIN*, *SITE*, and *DAMAGE*. Maximizing the log likelihood function (4.10), the maximum likelihood estimates of

$\beta = (\beta_1, \beta_2, \cdots, \beta_k)$ and σ were obtained (Table 4-5). A positive coefficient indicates that the corresponding variable has a positive impact on the mean of log survival time, and also on the survival probabilities. The likelihood ratio method was used to test the significance of the parameters. The likelihood ratio statistic and the p-value are also listed in the table.

The log-normal distribution was also fitted for spruce by maximizing the log likelihood function (4.11). It was found that the Weibull distribution model had a smaller log likelihood statistic. Therefore the Weibull distribution model, demonstrated as a better model, was chosen for spruce.

In the fitted spruce lifetime distribution model, the positive coefficient of height (0.0584274) shows that taller trees have higher survival probabilities. As a seedling grows taller, it gradually becomes large enough to be able to stand against numerous biotic and mechanical agents, it takes more advantages in competition with other seedlings or other vegetation, and has higher survival probability. The positive coefficient of $AWDEN^{-1}$ (94.5224759) indicates that the survival of spruce declines with increasing number of aspen trees. In aspen and spruce mixedwood stands, the shade intolerant aspen shows faster early growth and rapidly forms dominance on the site by occupying the upper layer of the canopy (Peterson and Peterson, 1992). With more aspen in the canopy, the more likely spruce trees are shaded, and the lower survival they have. The positive coefficient of good *SITE* (0.27317284) suggests that the seedlings survive better on better site. The positive coefficient of container *ORIGIN* (0.70986717), and negative coefficient of bareroot *ORIGIN* (-0.684178) show that the container planted spruce has the highest survival and the natural spruce has higher survival than the bareroot planted spruce. The *DAMAGE* for spruce was classified into 4 groups: suppression, damage from flooding or climatic extremes, other damage, and no damage. Among the four coefficients of the four classes, the coefficient of no damage (1.57663412) is the highest, which implies that all damages reduce the seedling survival. Damage from climatic extremes or flooding (1.14259358) reduces the

survival probability almost as much as other damage (1.10328981) except suppression. A suppressed white spruce seedling has the lowest survival probability. The estimated extreme value scale parameter $\hat{\sigma} = 1.03748254$, thus the Weibull scale parameter is obtained as $\hat{\delta} = \frac{1}{\hat{\sigma}} = 0.907024$. $\hat{\delta}$ is less than 1, which indicates the hazard rate is monotonously decreasing with the passage of time.

For lodgepole pine, it was assumed the lifetime follows a log-normal distribution. The normal distribution for the log lifetime was fitted by maximizing the log likelihood function (4.11) and the selected variables for equation (4.8) were: *AGE*, *HEIGHT*, $DENSITY^{-1}$, *DAMAGE*, *DW*, and *HS*. The parameter estimates were significant (Table 4-6). The coefficients of *AGE* (0.81789905) and *HEIGHT* (0.25457889) show that as a seedling grows older and taller, it gradually becomes more resistant to against numerous biotic and mechanical agents, and has higher survival probability. The positive coefficient of $DENSITY^{-1}$ (1049.49946) shows that stand density has a negative impact on lodgepole pine survival. In a seeded or natural pine stands, density is generally very high and trees grow relatively fast. More tree seedlings brings in more competition among them. The coefficients of poor drainage (-0.6858522) and deep humus depth (0.59014574) state that the survival of a pine is higher in well-drained stand with a deeper humus layer. In the pine model, the survival is significantly different between damaged trees (coefficient -0.3808868) and healthy trees, but is not significantly different among the trees damaged by the different agents. The reason might be that, as shown in Table 4-1, nearly 50% damage causes are unknown.

It was assumed that the lifetime for aspen follows a log-normal distribution. The normal distribution was fitted for the log lifetime of aspen. The selected variables for equation (4.8) were *AGE*, *HEIGHT*, $SWDEN^{-1}$, *DAMAGE*, *SITE*, and *DW*. The parameter estimates are listed in Table 4-7. The likelihood ratio method was used to test the significance of the parameters. The

likelihood ratio statistic and the p-value are also listed in the table. Same as in the spruce and pine model, both *AGE* (0.38349.41) and *HEIGHT* (0.00455328) have the positive impacts on survival. The survival is higher on good *SITE* (1.14113838), and lower with poor drainage (*DW*) (-1.1682132). The spruce density has a positive impact on aspen survival because of the negative coefficient of $SWDEN^{-1}$ (-398.02768). This is not surprising since the fast growing aspen is more competitive over spruce. More spruce may indicate better site productivity, which also positively impacts the survival of aspen. A healthy seedling (coefficient 2.44634809) has higher survival than a damaged one (2.06370673). A seedling damaged by herbicide treatment and suppression has the lowest survival.

4.4.4 Fitting evaluation

For the white spruce model, The residuals of the Weibull distribution model were computed based on equation (4.13). The ordered residuals were plotted against α_i (4.14) for uncensored data. The plot was approximately linear, which supports the assumption that the Weibull distribution fits the spruce lifetime distribution. For the aspen and lodgepole pine models, the residuals were calculated based on (4.15) and the procedure PROC UNIVARIATE on SAS (SAS Institute Inc. 1992) was used to check the normality for the uncensored residuals. The probability plot and W-test (SAS Institute Inc. 1992) showed that the log normal distribution assumption for the lifetime was acceptable.

In this study, another intuitive evaluation of the model was made based on the comparison of the predicted number with the actual number of trees. The evaluation was based on the same data set, since we could not find an independent data set with sufficient trees, and a prediction for a few independent trees is unable to show the general survival pattern. Based on the data, the trees which were alive, dead, and censored at each 2, 4, and 6 years were summarized and listed in Table 4-8. Excluding the censored trees, the number of live and dead

trees predicted from the parametric regression models (Tables 4-5, 4-6, and 4-7) were also listed to compare with the actual numbers. In Table 4-8, the predicted numbers of live trees are close to the observed frequencies, which indicates that the survival estimated from the models generally follows the actual survival pattern.

4.4.5 Model usage

For a given tree in a stand, inputting the values of explanatory variables, the parameters of the log lifetime distribution can be estimated based on the parametric regression model. Moreover the survival function, probability distribution function, and hazard function of lifetime can be estimated. With the survival function or probability distribution function, the survival probability in any time interval can be calculated. Given a tree list, the model can also predict the stand mortality. The following two examples show how to calculate the survival probability of an individual tree.

Example 1. In a white spruce plantation where $SITE = medium$, $DENSITY = 15000$ including 10000 aspen and 5000 white spruce, $DW = well - drained$, and $HS = deep$, a white spruce seedling, planted with bareroot method, is 2 year old and 10 cm in height. The seedling is healthy without any damage record.

According to the parameter estimates for white spruce in Table 4-5, the location parameter of the extreme value distribution is estimated as

$$\begin{aligned}\hat{\mu}(x) &= 1.6635172 + 0.0584274 \times 10 + 94.5334759 / 10000 - 2288.1016 / 15000 \\ &\quad -0.6841087 + 1.57663412 \\ &= 2.99723,\end{aligned}$$

and the scale parameter of the extreme value distribution $\hat{\sigma} = 1.10205069$. Since $\alpha = \exp(\mu(x))$ and $\delta = \frac{1}{\sigma}$ (Lawless 1982), $\hat{\alpha} = \exp(2.563189) = 20.02997$ and $\hat{\delta} = \frac{1}{\hat{\sigma}} = 0.907024$. We obtain

that the lifetime of the seedling follows the Weibull distribution with a shape parameter 20.02997 and a scale parameter 0.907024. Since the scale parameter is less than 1, its hazard function is decreasing as expected.

Based on the survival function of Weibull distribution in Table 4-3, the estimated survival function of the tree is $S(t)=\exp\left[-(t/20.02997)^{0.907024}\right]$. The current annual survival probability of the tree, $S(1)$, is equal to 0.936159, and the probability that the tree surviving longer than 5 years $S(5)=0.752762$. If the tree survives the current year, the probability that surviving another year is equal to the conditional probability $S(2)/S(1)=0.9439$. This conditional probability is similar to that provided by the logistic model. The survival function is plotted in Figure 4-2. As time passes, the survival function decreases.

If the tree is damaged by suppression, then the $\hat{\mu}(x)=1.420569$. The survival function becomes to $S(t)=\exp\left[-(1/4.139586)^{0.907024}\right]$. It is also plotted in Figure 4-2 as a comparison with the healthy one. The probability that the suppressed tree surviving longer than 5 years is only $S(5)=0.305188$.

Example 2. The usage of lodgepole pine model and aspen model are slightly different because the lifetime is a log-normal distribution instead of a Weibull distribution. In a seeded lodgepole stand with $DENSITY=15000$, $DW=poorly-drained$, and $HS=shallow$, a healthy lodgepole pine is one year old and 2 cm tall. Based on the parametric regression model (Table 4-6), the estimated location parameter for the normal distribution, a linear combination of the variables, is equal to $\hat{\mu}(x)=2.134994$. The scale parameter is $\hat{\sigma}=1.499197$. According to Table 4-3, the survival function of the tree can be obtained as $S(t)=1-\Phi\left[\left(\log(t)-2.134994\right)/1.499197\right]$. The function is plotted against time in Figure 4-3.

The probability that the tree surviving longer than 5 years is $S(5) = 0.641568$. If the stand is well-drained with a deep humus layer, the survival function is higher and also plotted in Figure 4-3. The probability that the tree survives longer than 5 year goes up to $S(5) = 0.893096$.

4.5 Summary and comments

Lifetime analysis method was used to analyze juvenile tree mortality for white spruce, lodgepole pine, and aspen in Alberta juvenile stands. A nonparametric analysis was used to portray the survival data, and provided a reasonable theoretical assumption of the lifetime distribution. A parametric regression model was developed relating the lifetime distribution to tree age, height, damage agents, regeneration method, stand density, and site productivity. The fitted model describes the white spruce, aspen, and lodgepole pine mortality trends, and allows survival projection to be made.

The methodology of the lifetime distribution analysis was systematically reviewed and applied to the juvenile tree mortality. No such a comprehensive analysis and discussion of the lifetime distribution model has been found in the previous tree mortality study. The parametric regression distribution model not only portrays the lifetime distribution for individuals, but also relates the tree and stand characteristics to the distribution. Compared with the logistic regression, it provides a broader view on the survival pattern and longer term survival prediction, and it is more sensitive to the damage agents. However, if the data do not follow any commonly used theoretical distribution, it will be much more difficult to fit. Also because of many censored observations, the extrapolation of the model beyond the data range could be dangerous. It is recommended that if the mortality prediction mainly focuses on the periodical survival probability, as commonly used in a stand growth model system, the logistic model is more applicable. If the survival pattern and the risk of mortality need to be viewed as time passes, and

the survival probability needs to be predicted for a long term, the lifetime distribution model is preferred.

The ideal data for the lifetime distribution study is the trees, with variety of age and size, in different stand conditions, that have been tracked until death. Then the fitted lifetime distribution model which covers a wide range of input variables can be used to describe and to predict the survival for the whole life span.

Table 4-1. Summary of possible mortality causes for white spruce, lodgepole pine, and aspen

Damage code	Probable cause of mortality	White spruce		Lodgepole pine		aspen	
		Mortality frequency	Mortality percentage	Mortality frequency	Mortality percentage	Mortality frequency	Mortality percentage
1	Insects	17	2.37	0	0.00	8	1.65
2	Disease	3	0.41	4	0.5	56	11.61
3	Rabbit browsing	1	0.13	0	0.00	0	0
4	Terminal insects	2	0.27	1	0.12	0	0
5	Browsing (other animals)	1	0.13	16	2.01	13	2.69
6	Fire	0	0.00	0	0.00	0	0
7	Mechanical	32	4.46	10	1.25	20	4.14
8	Windthrow	2	0.27	1	0.12	8	1.65
9	Climatic extremes	85	11.85	138	17.33	18	3.73
10	Flooding	137	19.10	29	3.64	21	4.35
11	Poor planting	31	4.32	0	0.00	0	0
12	Suppression	162	22.59	145	18.21	112	23.23
13	Frost heaving	17	2.37	2	0.25	1	0.18
14	Erosion	5	0.69	7	0.87	1	0.18
15	Unknown causes	162	22.59	391	49.12	56	11.61
16	Dead top	3	0.41	0	0.00	0	0.00
17	Poor seeded	43	5.99	52	6.53	1	0.2
18	Herbicide	0	0.00	0	0.00	169	35.06
19	Western gall rust	0	0.00	0	0.00	0	0.00
20	Armillaria root stock	1	0.13	0	0.00	0	0.00
21	Moldy planting stock	13	1.81	0	0.00	0	0.00
Σ		717	100.00	769	100.00	482	100.00

Table 4-2. Summary statistics of the lifetime data for white spruce, aspen, and lodgepole pine

White Spruce: *CENSOR*=0					
Variable	N	Mean	Std Dev	Minimum	Maximum
HEIGHT	717	7.758716	9.634323	1.000000	102.00000
AGE		1.859135	1.060181	1.000000	6.000000
DENSITY		11528.37	6889.87	1200.00	20000.00
SC_{SW}		0.547289	0.174324	0.057554	0.875000
LIFETIME		2.092050	0.990122	1.000000	5.000000
White Spruce: *CENSOR*=1					
Variable	N	Mean	Std Dev	Minimum	Maximum
HEIGHT	2936	9.262602	8.512432	1.000000	69.000000
AGE		1.994891	0.960714	1.000000	6.000000
DENSITY		12464.40	9830.43	950.00	42500.00
SC_{SW}		0.456452	0.225178	0.051344	1.000000
LIFETIME		5.277248	2.898411	1.000000	10.000000
Aspen: *CENSOR*=0					
Variable	N	Mean	Std Dev	Minimum	Maximum
HEIGHT	482	43.020746	28.149725	1.000000	142.000000
AGE		1.558091	0.548809	1.000000	4.000000
DENSITY		14210.35	5405.30	1550.00	20350.00
SC_{AW}		0.510524	0.195787	0.107569	0.837963
LIFETIME		2.041493	1.085885	1.000000	5.000000
Aspen: *CENSOR*=1					
Variable	N	Mean	Std Dev	Minimum	Maximum
HEIGHT	3654	53.719211	40.427471	2.000000	810.000000
AGE		1.676792	0.743121	1.000000	7.000000
DENSITY		13629.92	6700.68	950.00	42500.00
SC_{AW}		0.613574	0.210905	0.004405	0.909871
LIFETIME		5.613300	2.719508	1.000000	10.000000
Lodgepole pine: *CENSOR*=0					
Variable	N	Mean	Std Dev	Minimum	Maximum
HEIGHT	796	2.162060	0.993117	1.000000	8.000000
AGE		1.055276	0.228662	1.000000	2.000000
DENSITY		16414.39	6777.08	950.00	22800.00
SC_{PL}		0.925457	0.108313	0.042553	1.000000
LIFETIME		2.251256	0.825496	1.000000	5.000000
Lodgepole pine: *CENSOR*=1					
Variable	N	Mean	Std Dev	Minimum	Maximum
HEIGHT	3360	3.031845	2.211474	1.000000	30.000000
AGE		1.486011	0.701578	1.000000	3.000000
DENSITY		14636.97	8040.19	250.00	27850.00
SC_{PL}		0.888246	0.102268	0.010869	1.000000
LIFETIME		4.961011	2.644000	1.000000	11.000000

Table 4-3: The probability distribution function and the survival function for lifetime T

Lifetime Distribution	Distribution probability function: $f(t)$	Survival function: $S(t)$	Hazard function: $h(t)$
Weibull	$\left(\frac{\delta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\delta-1}\exp\left(-\left(\frac{t}{\delta}\right)^{\delta}\right)$	$\exp\left(-\left(\frac{t}{\alpha}\right)^{\delta}\right)$	$\frac{\delta}{\alpha}\left(\frac{t}{\alpha}\right)^{\delta-1}$
Log normal	$\frac{1}{(2\pi)^{1/2}\sigma t}\exp\left(-\frac{1}{2}\left(\frac{\log(t)-\mu}{\sigma}\right)^{2}\right)$	$1-\Phi\left(\frac{\log(t)-\mu}{\sigma}\right)$	$\frac{f(t)}{S(t)}$

Table 4-4. The probability distribution function and the survival function for log lifetime Y

T distribution	$Y=\log T$ Distribution	Probability distribution function for $\log T$: $f(y\|\boldsymbol{x})$	Survival function for $\log T$: $S(y\|\boldsymbol{x})$
Weibull	Extreme value	$\frac{1}{\sigma}\exp\left(\frac{y-\mu(\boldsymbol{x})}{\sigma}-\exp\left(\frac{y-\mu(\boldsymbol{x})}{\sigma}\right)\right)$	$\exp\left(-\exp\left(\frac{y-\mu(\boldsymbol{x})}{\sigma}\right)\right)$
Log-normal	Normal	$\frac{1}{(2\pi)^{1/2}\sigma}\exp\left(-\frac{1}{2}\left(\frac{y-\mu(\boldsymbol{x})}{\sigma}\right)^{2}\right)$	$1-\Phi\left(\frac{y-\mu(\boldsymbol{x})}{\sigma}\right)$

Table 4-5. Parameter estimates of the extreme value log-lifetime distribution for white spruce

Variable	DF	Estimate	Std Err	Chi-square	Pr>Chi	Label
INTERCEPT	1	1.663517	0.420782	15.62936	0.0001	Intercept
HEIGHT	1	0.058427	0.013509	18.70645	0.0001	
AWDEN[1]	1	94.533476	26.74026	12.49507	0.0004	
DENSITY[1]	1	-2288.1016	501.471	20.81896	0.0001	
ORIGIN	2			47.05673	0.0001	
	1	-0.684108	0.255055	7.194166	0.0073	Bareroot planting
	1	0.7098672	0.201894	12.36248	0.0004	Container planting
	0	0	0			Ingress
DAMAGE	3			22.39332	0.0001	
	1	1.1425936	0.544975	4.395728	0.0360	Climatic extremes or Flooding
	1	1.1032898	0.443865	6.178415	0.0129	Other damage
	1	1.5766341	0.428811	13.90494	0.0002	Health (no damage)
	0	0	0			Suppression
SITE	1			5.223399	0.0223	
	1	0.2731728	0.119526	5.223399	0.0223	Good site
	0	0	0			Medium site
SCALE	1	1.1020507	0.03941			Extreme value scale

Table 4-6. Parameter estimates of the normal log-lifetime distribution for lodgepole pine

Variable	DF	Estimate	Std Err	Chi-square	Pr>Chi	Label
INTERCEPT	1	1.4238559	0.144979	96.455	0.0001	Intercept
AGE	1	0.8178991	0.122828	44.34061	0.0001	
HEIGHT	1	0.2545789	0.036604	48.37029	0.0001	
DENSITY[1]	1	1049.4995	276.7773	14.37817	0.0001	
DAMAGE	1			6.821425	0.0090	
	1	-0.380887	0.145834	6.821425	0.0090	Damage
	0	0	0			Health
DW	1			69.56501	0.0001	
	1	-0.685852	0.082231	69.56501	0.0001	Poorly-drained
	0	0	0			Well-drained
HS	1			44.5884	0.0001	
	1	0.5901457	0.088379	44.5884	0.0001	Deep (≥ 16 cm)
	0	0	0			Shallow (< 16 cm)
SCALE	1	1.4491966	0.042525			Normal scale parameter

Table 4-7. Parameter estimates of the normal log-lifetime distribution for aspen

Variable	DF	Estimate	Std Err	Chi-square	Pr>Chi	Label
INTERCEPT	1	1.3301947	0.251316	28.01487	0.0001	Intercept
AGE	1	0.3834904	0.100875	14.45238	0.0001	
HEIGHT	1	0.0045533	0.001950	5.454608	0.0195	
SWDEN^{-1}	1	-398.0277	161.9906	6.037358	0.0140	
DAMAGE	2			206.4317	0.0001	
	1	2.0637067	0.196686	110.0902	0.0001	Damage
	1	2.4463481	0.170281	206.3963	0.0001	Health
	0	0	0			Suppression or herbicide
SITE	1			20.59177	0.0001	
	1	1.1411384	0.251473	20.59177	0.0001	Good site
	0	0	0			Medium site
DW	1			88.99417	0.0001	
	1	-1.168213	0.123834	88.99417	0.0001	Poorly-drained
	0	0	0			Well-drained
SCALE	1	1.7342575	0.070988			Normal scale parameter

Table 4-8. Comparison between the actual and predicted number of trees

Species	Prediction interval	Actual number of live trees	Predicted number of live trees	Actual number of dead trees	Predicted number of dead trees	Number of censored trees
White spruce	2-year	2634	2566	172	240	847
	4-year	1917	2115	580	382	1156
	6-year	1310	1512	621	419	1722
Pine	2-year	4080	3867	72	285	33
	4-year	2183	2354	674	502	1328
	6-year	1418	1614	796	663	1908
Aspen	2-year	3366	3380	195	181	575
	4-year	2943	3056	422	309	771
	6-year	1920	2032	458	346	1758

Figure 4-1. Empirical hazard functions by species

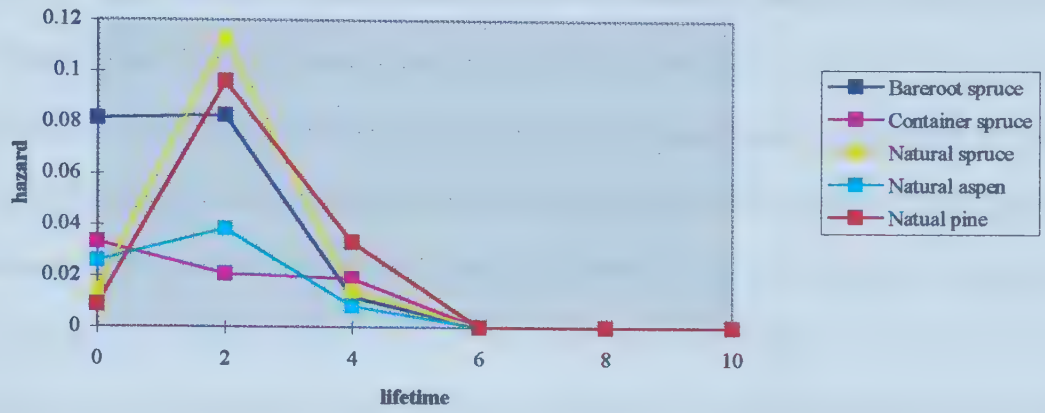


Figure 4-2. Survival functions for white spruce (Example 1)

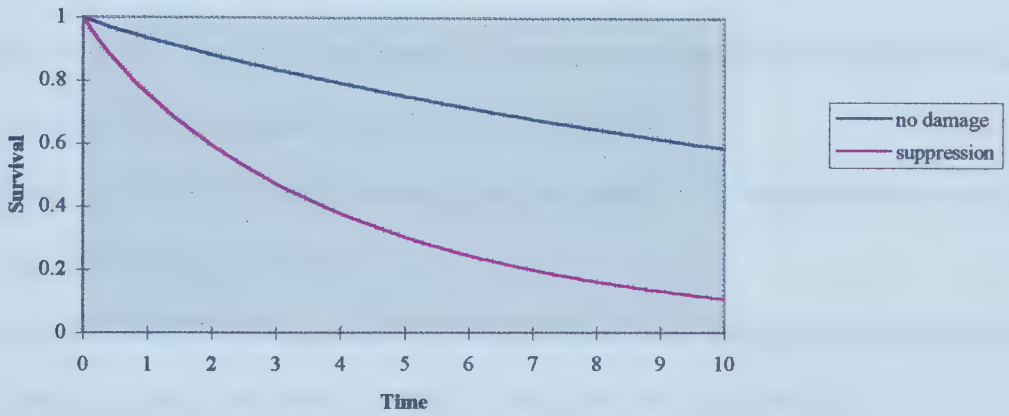
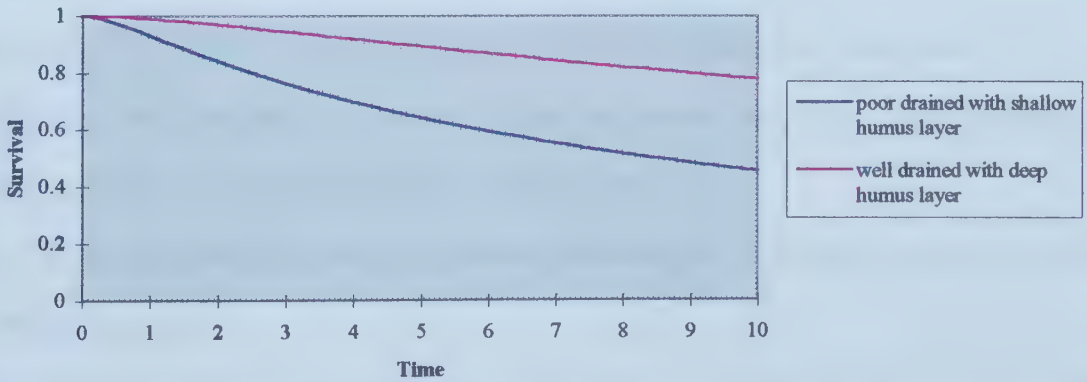


Figure 4-3. Survival functions for lodgepole pine (Example 2)



4.6 References

Barbour, M.G., Pavlik, B.M. and Antos, J.A. 1990. Seedling growth and survival of red and white fir in a Nevada ecotone. Amer. J. Bot. 77(7): 927-938.

Brand, D. G. 1991. The establishment of boreal and sub-boreal conifer plantations: an integrated analysis of environmental conditions and seedling growth. Forest Science. 37(1):68-100.

Buchman, R.G., Pederson, S.P., and Walters, N.R. 1983. A tree survival model with application to species of the Great Lakes region. Can. J. For. Res., 13:601-608.

Gilday, D.W. 1990. Monitor Plot survey manual. Published by Alberta Land and Forest Service.

Hamilton, D.A. 1974. Event probabilities estimated by regression. USDA Forest Service research paper. INT-152, 18p.

Hamilton, D.A. 1986. A logistic model of mortality in thinned and unthinned mixed conifer stands of Northern Idaho. Forest Science, 32:989-1000.

Hamilton, D.A. 1990. Extending the range of applicability of an individual tree mortality model. Can. J. For. Res., 20:1212-1218.

Hann, D.W. 1980. Development and evaluation of an even- and uneven-aged ponderosa pine / Arizona fescue stand simulator. U.S.D.A. For. Serv. Res. Pap. INT-267.

Harcombe, P.A. 1987. Tree life table - simple birth, growth, and data encapsulate life histories and ecological roles. BioScience Vol. 37 No. 8.

Kalbfleisch, J. D. 1980. The statistical analysis of failure time data. New York: Wiley.

Lawless, J.F. 1982. Statistical models and methods for lifetime data. Wiley series in probability and mathematical statistics. John Wiley & Sons. New York.

Lee, E.T. 1992. Statistical method for survival data analysis. 2nd edition. New York . John Willey & Sons, Inc.

Maguire, D.A., and Maguire, C.C. 1987. Analyzing treatment effects on Douglas-fir seedling survival with an extreme value regression model. Can. J. For. Res. 17:1267-1630.

Monserud, R.A. 1976. Simulation of forest tree mortality. Forest Science, 22:438-444.

Moser, J.W. 1972. Dynamics of an uneven-aged forest stand. Forest Science, 18:184-191.

Nelson, W. 1990. Accelerated testing: statistical models, test plans and data analysis. New York: Wiley.

Peterson E.B. and Peterson, N.M. 1992. Ecology, management, and use of aspen and balsam poplar in the prairie provinces. Forestry Canada.

SAS Institute Inc. 1992. SAS user's guide: statistics. Version 6 ed., Cary, NC 27512-8000.

Shiver, B.D., Borders, B.E. Page, H.H. Jr., and Raper, S.M. 1990. Effect of some seedling morphology and planting quality variables on seedling survival in Georgia Piedmont. South. J. Appl. For. 14(3):109-114.

Somers, G.L., Oderwald, R.G., Harms, W.R., and Langdon, O.G. 1980. Predicting mortality with a Weibull distribution. Forest Science, 26:291-300.

Turner, I.M. 1990. Tree seedling growth and survival in a Malaysian rain forest. Biotropica. 22(2):146-154.

Waters, W.E., DeMars, C.J. Jr., and Cobb, F.W. Jr. 1991. Analysis of early mortality of Douglas-fir seedling in postharvest plantings in northwestern California. Forest Science. 37(3):802-826.

Chapter 5

Juvenile tree height versus age relationships

5.1 Introduction

During the last several decades, extensive natural stands have been harvested and replaced by regenerated stands in Alberta. As the forest industry continues to grow and expand, the knowledge and management for these juvenile stands have become increasingly important for predicting yield in regenerated stands. The focus of the growth and yield modeling in this province has been shifted towards the regenerated stands in the past several years, and a regenerated stand growth model is needed.

A juvenile tree height prediction model is an essential component in a regenerated stand growth model. The estimation of tree volume, as well as the description of the regenerated stand development, relies heavily on accurate height prediction. The height versus age relationships predict cumulative height growth for individual trees based on age and other variables.

Since many growth relationships for older trees are based on variables other than age (Huang and Titus 1994, Burkhart *et al.* 1972, Curtis at el. 1981, Wykoff *et al.* 1982), and yield tables include age, the juvenile stand model allows linking the age variable needed when producing yield tables with a stand growth model that may not require age. The juvenile stand model is also useful in simulating the tree height until it reaches a 1.3 metre height. At that point diameter at breast height is available and can be used in other tree growth relationships (Huang and Titus 1994, Arabatzis and Burkhart 1992, Wang and Hann 1988).

Because the interest in regenerated stands is very recent, remeasured growth data are not commonly available for stands younger than about 30-40 years. Most natural stand permanent sample plots were established decades ago in older stands and it is only within the last 15 years

that permanent sample plots and other research projects have been established in juvenile stands (Gilday 1990, Alberta Forest Service 1993).

This study evaluated juvenile individual tree height and age relationships for three major species in Alberta regenerated pure and mixedwood stands: trembling aspen (*Populus tremuloides* Michx), white spruce (*Picea glauca* (Moench) Voss), and lodgepole pine (*Pinus contorta* var. *latifolia* Engelm). The data were permanent sample plots from regenerated stands younger than 15 years, on which most individual trees were measured for both age and height. The fitted height versus age relationships predict total tree height given tree age, regeneration origin, damage code, stand density, species composition, and site productivity class. The curve-fitting method was to first select an appropriate height-age equation as the base function, and then incorporate other tree and stand level variables such as tree regeneration origin, damage indicator, stand density, species composition, and site productivity to predict the parameters of the base function, using the parameter prediction method (Clutter *et al.* 1983). The height growth model was fitted using nonlinear least squares regression. To account for the unequal error variance found from analysis of residuals, a weighting factor was applied to achieve the uniform error variance. To minimize possibility of the serial correlation in the data, only one remeasurement of the series was used to fit the model. The remaining data was used to test the model performance.

5.2 Materials and methods

5.2.1 Data

The Alberta Land and Forest Services provided the data for this study. The data are permanent sample plots designed to "monitor" juvenile stand establishment and growth on cut-blocks (Gilday 1990). The plots were established in various ecoregions across the province since 1980. There were 107 plots that were remeasured every two years and with same plots remeasured

up to 7 times. At plot establishment, harvest, scarification, regeneration treatment, ecoregion, moisture regime, drainage condition, duff depth and some other stand characteristics were recorded for each plot. The regeneration treatments included left-for-natural, seeded, and planted with different species. Of the 107 plots, 69 were planted with white spruce in which 33 were planted with 1-year-old containerized seedling and 36 were planted with 3-year-old bareroot seedlings; 22 were left for natural or seeded with lodgepole pine or mixed species; 7 were planted with lodgepole pine; 3 were seeded with white spruce; and 6 were seeded with forage and unspecified species. The last three groups were not of sufficient size for analysis and 6 plots were re-seeded or thinned. They were excluded from this study.

Each plot consisted of 40 quadrats (2 m × 2.5 m) that were alternately established on each side of four 25 metre transects distributed over a cut-block (Gilday 1990). Lodgepole pine, jack pine, white spruce, black spruce, and all fir species were defined as acceptable species. In a quadrat, each tree of acceptable species and the three tallest deciduous trees were tagged and measured. Measurements included species, age, height, regeneration origin, and damage code. Other deciduous trees were counted by species. Tree age was estimated when a tree was first found on the site. The tree regeneration origin indicated whether the tree was planted or seeded. The damage code indicated the damage categories for live trees and reported possible causes of mortality when trees were first found dead (Gilday 1990). It included 21 categories such as suppression, flooding, climatic extremes, diseases, insects, and etc.

The data consisted of repeated height and age measurements on the same trees over an extended period. The repeated measurements tended to be correlated, and did not satisfy the assumption of independence. To remove the serial correlation among repeated measurements of the same trees, only one measurement was used. Since there were new plots established at most remeasurement times, the last measurements of all the plots, covering both oldest plots and youngest plots, were chosen to fit the height versus age relationships. In total there were 2648

white spruce, 2678 aspen, and 3051 lodgepole pine trees. The remaining remeasurements were used in model evaluation since there was no independent data set providing the same input variables and representing the population on which the model was to be used.

5.2.2 Explanatory variables

Tree age (*AGE*) was the major explanatory variable in this study. Increasing tree height was expected with increasing age. However age alone is not sufficient to predict height since many other factors are at work.

The juvenile trees were vulnerable to many damage agents. Checking the damage code for the data, of total trees, 30 percent of spruce, 60 percent of aspen, and 16 percent of lodgepole pine were damaged by suppression, flooding, climatic extremes, diseases, insects, and etc. Those damaged trees might show less height growth and lower survival probability. In this study, a health indicator, H, was given to each tree. If a tree was not damaged, then H=1, otherwise H=0.

Tree regeneration origin had three classes: bareroot for trees that were nursery grown bareroot seedlings, container for trees that were nursery grown containerized seedlings, and ingress for trees that were germinated from seed sources on the site. Since all blocks were harvested by clear-cutting, a small number of advance trees, those which had grown on the site for at least one year prior to harvesting, were excluded from this study. *BAREROOT* and *CONTAINER* were two indicator variables. *BAREROOT*=1 if the seedling was a bareroot seedling, otherwise it was equal to zero. *CONTAINER*=1 if the seedling was a containerized seedling, otherwise zero. For ingress seedlings, both *BAREROOT* and *CONTAINER* were equal to zero.

Although only a few of the deciduous trees were tagged and remeasured, all of them were counted by species. This made it possible to calculate total density and species composition as additional variables. The total density (*DENSITY*) was defined as the total number of trees per

ha for all species. The species composition (SC_{SP}) was equal to the number of trees per ha for target species divided by number of trees per ha for all species. SC_{SW}, SC_{AW}, and SC_{PL} denoted white spruce composition, aspen composition, and lodgepole pine composition.

The site productivity classification of the cut-blocks was identified from the Phase 3 Inventory. Since almost all the cut-blocks were located in good and medium sites, the site class only had two categories: good and medium. *SITE*=1 where the site class was good, and *SITE*=0 where the site class was medium. In addition, the drainage condition, an assessment of soil drainage characteristics of each block was also considered. In the original data, the drainage condition had 12 classes. In this study only two classes were considered, well-drained or poorly-drained. *DW*=1 if the stand was well-drained, and *DW*=0 if the stand was poorly-drained.

5.2.3 Model development

To develop height growth relationships, an appropriate height-age base curve was first selected by a preliminary analysis of the data, and then other tree and stand level variables *BAREROOT*, *CONTAINER*, *H*, *DENSITY*, SC_{SP}, *SITE*, and *DW* were combined to form the parameters of the base function using the parameter prediction method (Clutter *et al.* 1983).

Two candidates for the height-age base function were considered in this study. The first one was an exponential curve:

$$HEIGHT = \alpha\left[\exp\left(\beta AGE\right) - 1\right], \tag{5.1}$$

where *HEIGHT* is a tree height (cm), *AGE* is the tree age, and α and β are unknown parameters. The curve starts from zero and montonically increases with increasing age. The selection of function (5.1) was based on an examination of the height and age relationships as revealed by plotting total tree height against age for the three species. The second candidate was the commonly used Chapman-Richards function (Lundgren and Dolid 1970, Ek 1971, Payandeh 1974, Alemdag 1991):

$$HEIGHT = a(1 - e^{-bAGE})^c \quad (5.2)$$

where *HEIGHT* is total tree height, *AGE* is the tree age, and *a*, *b*, and *c* are unknown parameters. It is a sigmoidal curve, with an inflection point occurring in the younger age and a maximum value *a* indicating the potential maximum growth.

Both models were fitted to the data, and the more appropriate one was chosen based on the goodness-of-fit and the application consideration. After that, other tree and stand level variables *BAREROOT*, *CONTAINER*, *H*, *DENSITY*, SC_{SP}, *SITE*, and *DW* were combined to form the parameters of the base function using the parameter prediction method (Clutter *et al.* 1983).

The fitting was accomplished by the nonlinear least squares regression utilizing the procedure PROC NLIN on SAS/STAT software (SAS Institute Inc. 1992). The Marquardt iterative method (Marquardt 1963) was applied to solve the nonlinear least squares equations. To ensure the solution was global rather than local, multiple starting values for parameter estimates were provided. A weighting factor was included to achieve a uniform error variance of residuals. The goodness-of-fit was evaluated using the mean squared error (MSE), the coefficient of determination (R^2), and plots of residuals. Prediction performance was evaluated using the other remeasurement data.

5.3 Results

5.3.1 Preliminary model fits

A total number of 2648 white spruce, 2678 aspen, and 3051 lodgepole pine trees were used to fit the model. The summary statistics are listed in Table 5-1. All aspen and lodgepole pine trees were 1 to 12 years old, and white spruce trees were 1 to 14 years old.

In a preliminary analysis, both base functions (5.1) and (5.2) were fitted to the data. It was found that the fitted exponential curve (5.1) was closer to the data, with a lower mean squared error (MSE) and a higher coefficient of determination (R^2). The preliminary analysis also

found that both models gave poor predictions of height beyond the data range. The exponential function (5.1) approached infinity rapidly with increasing age, which indicated that the curve overestimated height growth for older trees. The Chapman-Richards function underestimated tree height beyond the data range by presenting the maximum height based on the data as the potential maximum height growth.

The juvenile height growth relationships in this study were intended for trees which were less than 1.3 metre in height. Once the trees passed 1.3 metre in height, other growth relationships (Huang and Titus 1994), with diameter at breast height as an input variable, were planned to replace the height and age relationships for height prediction. Based on this design and the fact that equation (5.1) had a better fit to the data, the truncated exponential curve of (5.1):

$$HEIGHT = \alpha\left[\exp(\beta AGE) - 1\right], \text{ for } HEIGHT \le 1.3 \text{ m}, \tag{5.3}$$

was chosen as the base function of the height growth model, and the usage of the model was limited to the trees less than 1.3 metre in height.

The explanatory variables *BAREROOT*, *CONTAINER*, *H*, *DENSITY*, SC_{SP}, and *SITE* were incorporated into the parameters α and β of the base function (5.3) by using the parameter prediction method (Clutter *et al.* 1983). The final model for white spruce was:

$$\begin{aligned} HEIGHT &= \alpha\left[\exp(\beta AGE) - 1 + \gamma_1 CONTAINER\right] \\ \alpha &= \alpha_1 \exp(\alpha_2 BAREROOT + \alpha_3 CONTAINER + \alpha_4 H + \alpha_5 DENSITY + \alpha_6 SC_{SW} + \alpha_7 DW), \\ \beta &= \beta_1 \end{aligned} \tag{5.4}$$

for $HEIGHT \le 1.3$ m. Note that model (5.4) has a non-zero intercept term for container planted spruce. It violated the restriction that $HEIGHT=0$ when $AGE=0$. The reason to accept the intercept term was that the model with zero intercept was found to be underestimating the height for 1 to 3-year-old container planted spruce. When younger than 1 year, the containerized seedlings were on nursery beds and grew very fast. The acceleration of height growth was much

higher in the first year than after planting. The change of the acceleration was continuous, but might not be smooth. The model with zero intercept at age zero could not express the growth acceleration for the first year on the nursery beds, and underestimated the growth for the first few years. Therefore to make a better fit, the height prediction for age around zero for the container planted trees was sacrificed, and the non-zero intercept was included.

For aspen and lodgepole pine, there was no problem with zero intercept, since they were all ingress seedlings. The model for aspen was found

$$\begin{aligned} HEIGHT &= \alpha\left[\exp(\beta AGE) - 1\right] \\ \alpha &= \alpha_1 \exp\left(\alpha_2 H + \alpha_3 DENSITY + \alpha_4 SC_{AW} + \alpha_5 SITE\right), \\ \beta &= \beta_1 \end{aligned} \tag{5.5}$$

for $HEIGHT \leq 1.3$ m, and the model for lodgepole pine was:

$$\begin{aligned} HEIGHT &= \alpha\left[\exp(\beta AGE) - 1\right] \\ \alpha &= \alpha_1 \exp\left(\alpha_2 H + \alpha_3 DENSITY + \alpha_4 SC_{PL} + \alpha_5 SITE + \alpha_6 DW\right), \\ \beta &= \beta_1 \end{aligned} \tag{5.6}$$

for $HEIGHT \leq 1.3$ m.

5.3.2 *Final fits*

The nonlinear least squares fits of the height prediction model (5.4), (5.5), and (5.6) were accomplished using the procedure PROC NLIN on SAS/STAT software (SAS Institute Inc. 1992). By examining the plot of the studentized residual against predicted total tree height, it was found that the variance of the residual increased with increasing the predicted value. To achieve homogeneous error variance, a weighted nonlinear least squares technique was applied. Several functions of *AGE* were used as the weighting factor, and the plot of the studentized residual against the predicted tree height was examined in each case. It was found the model with

weighting factor $w = 1/AGE^2$ produced estimates with satisfactory residual plots. The residual plot for white spruce is shown in Figure 5-1.

The final fits for the height growth models (5.4), (5.5), and (5.6) are presented in Tables 5-2, 5-3, and 5-4 for white spruce, aspen, and lodgepole pine. The mean squared error (MSE) and the coefficient of determination (R^2) were also calculated and listed. The coefficients of determination (R^2) for white spruce, aspen, and lodgepole pine are 0.7448, 0.6306, and 0.7653 respectively, which are not as high as those commonly found for mature tree height and age relationships. This indicates that the juvenile tree height growth was quite variable. The variability probably was due to genetic and microsite differences, and other factors which could not be presented in the model.

5.3.3 Model testing

The remaining remeasurements of the data, not used to fit the model, were used for testing the model performance, since there was no independent data set available with age on each individual tree. The testing data included 2657 white spruce trees, 2387 aspen trees, and 3038 lodgepole pine trees. For the three species, the actual height from the testing data was compared with the predicted height by models (5.4), (5.5), and (5.6) using the estimated coefficients in Tables 5-2, 5-3, or 5-4. The difference was obtained by subtracting the predicted height from the actual height. The mean and the standard deviation of the difference were computed and listed in Table 5-5 for each species. A *t*-test of the null hypothesis that the mean of the difference was zero was conducted according to the method described by Rawling (1988). The *t* statistics were equal to 0.912891, -1.152981, 0.504288. The *p*-values for the three *t*-tests were 0.3614, 0.2490, and 0.6141. There was no evidence to against the null hypothesis at level $\alpha = 0.05$. This infers that the prediction was an unbiased estimate of the actual height. The

differences were also plotted against *AGE*, and it was found that although the variance increased with increasing age, it was relatively small and acceptable. The model performed well.

5.4 Discussion

The juvenile height growth relationships (5.4), (5.5), and (5.6) predict total heights for individual trees given tree age, tree regeneration origin, damage indicator, stand density, species composition, and site productivity. These models were designed to predict height for trees that have not reached 1.3 m in height. Using the fitted models for older stands will overestimate tree heights.

For white spruce, the model shows that tree heights are significantly different among the tree regeneration origins. At any given age, if other variables remain the same, container planted spruce is taller than the bareroot planted spruce (Figure 5-2). Both container seedlings and bareroot seedlings are taller than the natural regeneration.

For all three species, the damage indicator is also a significant factor affecting height. As mentioned in the section of materials and methods, 30% of spruce, 60% of aspen, and 16% of lodgepole pine in the data were damaged by suppression, flooding, climatic extremes, diseases, insects, and etc. The positive coefficients of the variable *H*, which are 0.225894 for white spruce, 0.133847 for aspen, and 0.164135 for lodgepole pine, indicate that trees without any damage have higher height predictions when other variables are the same.

Although the data did not support the study of competition between trees and other vegetation, the negative coefficients of stand density show that the competition among trees has a significant negative impact on height growth. This confirms the general expectation that greater density in young stands leads either to heavy juvenile mortality or severe growth stagnation (Jones and Trujillo 1975), although in mature stands density may positively stimulate the height growth (Huang and Titus 1994).

Site productivity has a positive impact on height growth, as indicated by the positive coefficient of site productivity (*SITE*) and the drainage class (*DW*). This is a generally acknowledged effect, and is compatible with many studies on height growth (Ek1971, Payandeh 1974, Alemdag 1991).

In spruce and aspen mixedwood stands, after harvesting, reproduction by root sucking assists the shade intolerant aspen to achieve fast early growth and domination of the site by occupying the upper layer of the canopy (Mueggler 1989). Therefore aspen has the competitive advantage over shade tolerant white spruce that exhibits slow juvenile growth (Mueggler 1989). This does not mean that aspen has a negative impact on spruce height growth. In the fitted model, the negative coefficient of species composition in the model for spruce (-0.766986) shows that in the white spruce plantations with lower spruce composition and higher deciduous composition, white spruce grows taller. Similar conclusions were found in other studies. Russell (1963) found that white spruce planted under a nurse canopy of aspen providing about 30 percent shade for the first 10 to 12 years were usually taller than those in the open. Olson and Perala (1981) corroborated that shading of 1 to 4 year old white spruce seedlings had a positive effect on height growth. They also found that overhead cover helped protect the seedlings from attacks by the yellowheaded spruce sawfly (*Pikonema alaskensis*) without affecting height growth.

In the aspen model (5.5), the positive coefficient of aspen composition (0.645743) shows that when stand density and site productivity remain the same, aspen composition has a positive influence on the height growth, that is, aspen grows taller in stands with more aspen trees. This may reflect Jones and Trujillo's (1975) comment that after clear-cutting of the mixedwood stands, new suckering increased aspen stocking for several years, and the aspen suckers grew very fast.

When stand density and site productivity remain the same, lodgepole pine composition has a positive influence on the height growth, which is presented by the positive coefficient of

the lodgepole pine composition (0.623597) in model (5.6). Lodgepole pine is a shade intolerant species, and it grows better in pure stands than in mixedwood stands.

Figure 5-2 shows the predicted height growth for all species less than 1.3 metre in height in the regenerated stands with medium site and density 15,000 per ha. Aspen grows much faster than coniferous seedlings. Container spruce seedlings achieve better early height growth than bareroot seedlings. Both container and bareroot spruce seedlings are taller than natural spruce regeneration. For natural lodgepole pine, at very young ages, the height is greater than the height of natural spruce, but less than the height of planted spruce. Later on, the height growth of pine will exceed planted spruce and show a high potential increase.

In Figure 5-2, the height predictions for aspen, lodgepole pine, and planted white spruce trees fall in the data range used to build the models (Table 5-1). Only the height prediction for natural spruce surpasses the maximum age 14 of the data. The extrapolation of the growth curve for natural spruce to 20 years shows the height growth as expected from general empirical knowledge of the juvenile stands, and is necessary since the intent is to use this relation to simulate height growth until the tree reaches 1.3 m. Other extrapolation is dangerous and should be avoided. Better height growth relations for juvenile stands will be possible when growth data is available for stands up to about 30 years.

In summary, the height versus age relationships were evaluated for the major species in Alberta mixedwood stands. A truncated exponential function was used as a height-age base function, and the parameter prediction method was used to incorporate additional tree and stand level variables into the parameters of the based function. The model predicts cumulative height growth for individual trees based on tree age, regeneration origin, damage indicator, stand density, species composition, and site productivity.

Table 5-1: Summary statistics for juvenile tree and stand characteristics by species

Species	Variable	Mean	Std. Dev.	Minimum	Maximum
White spruce	*HEIGHT*	33.1567	32.1466	1.0000	263.0000
	AGE	6.7926	3.1453	1.0000	14.0000
	DENSITY	1.8500×10^4	0.9968×10^4	0.0450×10^4	3.8700×10^4
	SC_{SW}	0.2720	0.1567	0.02571	0.8421
Aspen	*HEIGHT*	142.4037	148.9592	1.0000	999.0000
	AGE	5.1527	3.1708	1.0000	12.0000
	DENSITY	1.5968×10^4	0.6458×10^4	0.0950×10^4	3.8700×10^4
	SC_{AW}	0.6692	0.2067	0.0044	0.9237
Lodgepole pine	*HEIGHT*	38.9469	45.2975	1.0000	305.0000
	AGE	5.8446	3.2185	1.0000	12.0000
	DENSITY	1.7900×10^4	0.8475×10^4	0.0850×10^4	3.4550×10^4
	SC_{PL}	0.8118	0.1757	0.0163	0.9950

Table 5-2. Parameter estimates for white spruce model (5.4)

Parameter	Estimate	Asymptotic Std. Error	Asymptotic 95% Confidential Interval Lower	Upper
α_1	22.167321	2.061605	18.124721	26.209922
α_2	0.710512	0.033496	0.644829	0.776195
α_3	0.607727	0.040315	0.528673	0.686782
α_4	0.225849	0.018162	0.1902339	0.261465
α_5	-0.068565×10^{-4}	0.011949×10^{-4}	-0.091996×10^{-4}	-0.045134×10^{-4}
α_6	-0.766986	0.072396	-0.908947	-0.625026
α_7	0.096612	0.017325	0.062640	0.130583
β_1	0.096272	0.006118	0.084276	0.108269
γ_1	0.333583	0.043880	0.247539	0.419628

Table 5-3. Parameter estimates for aspen model (5.5)

Parameter	Estimate	Asymptotic Std. Error	Asymptotic 95% Confidential Interval Lower	Upper
α_1	122.295180	13.031940	96.740985	147.849375
α_2	0.133847	0.024044	0.086698	0.180995
α_3	-0.116690×10^{-4}	0.018596×10^{-4}	-0.153154×10^{-4}	-0.080226×10^{-4}
α_4	0.645743	0.066015	0.516296	0.775191
α_5	0.099123	0.024836	0.050422	0.147825
β_1	0.111611	0.006264	0.099328	0.123894

Table 5-4. Parameter estimates for lodgepole pine model (5.6)

Parameter	Estimate	Asymptotic Std. Error	Asymptotic 95% Confidential Interval Lower	Upper
α_1	1.385364	0.196329	1.000405	1.770322
α_2	0.164135	0.024779	0.115549	0.212721
α_3	-0.039035×10^{-4}	0.01579×10^{-4}	-0.069189×10^{-4}	-0.008880×10^{-4}
α_4	0.623597	0.082388	0.462060	0.785133
α_5	0.243658	0.050609	0.144426	0.342891
α_6	0.626597	0.044140	0.540048	0.713146
β_1	0.355587	0.007829	0.340234	0.370938

Table 5-5. The t-test statistics for the model testing

Species	N	Mean	Std. Error	T	Prob>|T|
White spruce	2657	0.240066	0.262973	0.912891	0.3614
Aspen	2387	-2.099269	1.820733	-1.152981	0.2490
Lodgepole pine	3038	0.239757	0.475439	0.504288	0.6141

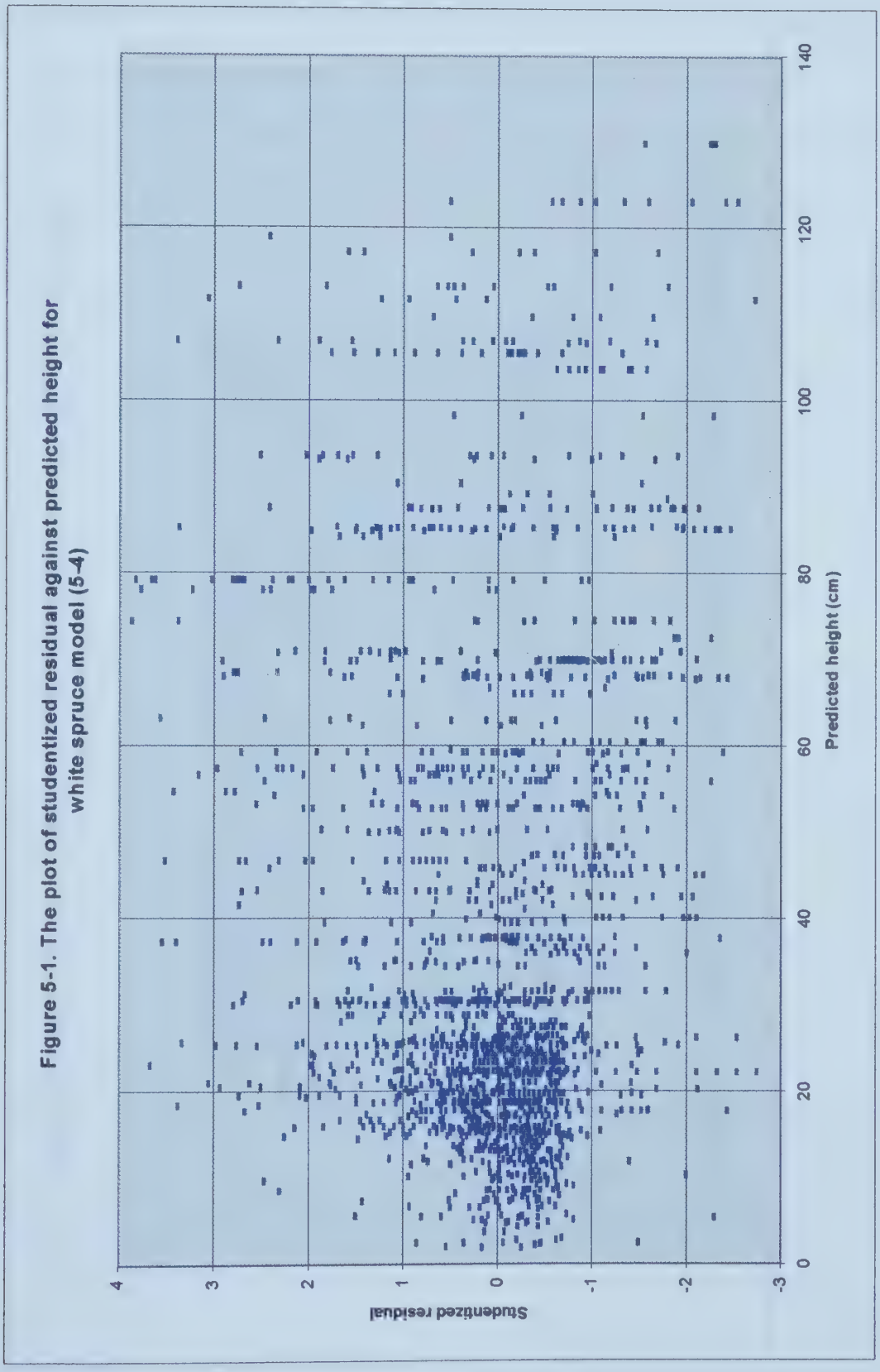
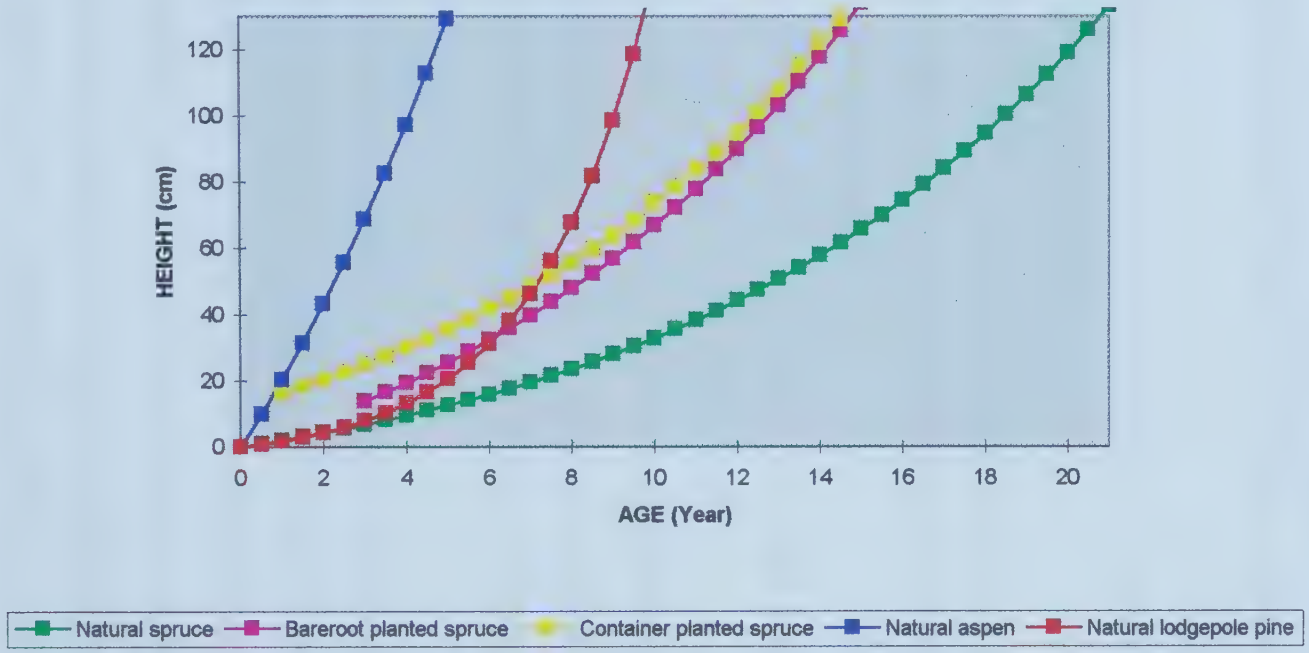


Figure 5-1. The plot of studentized residual against predicted height for white spruce model (5-4)

Figure 5-2. Height versus age relationships for healthy seedlings given *SITE*=medium and *DENSITY*=15000



5.5 References

Alemdag, I.S. 1991. National site-index and height-growth curves for white spruce growing in national stands in Canada. Can. J. For. Res. 21: 1466-1474.

Arabatzis , A.A., and Burkhart, H.E. 1992. An evaluation of sampling methods and model forms for estimating height-diameter relationships in loblolly pine plantations. Forest Science, 38: 192-198.

Alberta Forest Service 1993. Stand dynamics system field remeasurement manual. Timber Management Branch, Alberta Forest Service.

Burkhart, H.E., Parker, R.C., Strub, M.R., and Oderwald, R.G. 1972. Yield of old-field loblolly pine plantations, School of Forestry and Wildlife Resources, Virginia Polytechnic Institute and State University, Publication FWS-3-72.

Clutter, J.L., Fortson, J.C., Pienaar, L.V., Brister, G.H., and Bailey, R.L. 1983. Timber management: a quantitative approach. John wiley & sons.

Curtis, R.O., Clendenen, G.W., and DeMars, D.J. 1981. A new stand simulator for coast Douglas-fir:DFSIM user's guide. USDA For. Serv. Gen. Tech. Rep. PNW-128.

Ek, A.R. 1971. A formula for white spruce site index curves. University of Wisconsin, Madison. For. Res. Note No. 161.

Gilday, D.W. 1990. Monitor Plot survey manual. Published by Alberta Land and Forest Service.

Huang, S. and Titus, S.J. 1994. An age-independent individual tree height prediction model for boreal spruce-aspen stands in Alberta. Can. J. For. Res., 24(7): 1295-1301.

Jones, J. R. and Trujillo, D.P. 1975. Development of some young aspen stands in Arizona. USDA rocky mountain forest and range experiment station. Research paper RM 151. 11p.

Lundgren, A.L., and Dolid, W.A. 1970. Biological growth functions describe published site index curves for Lake States timber species. U.S. For. Serv. Res. Pap. NC-36.

Marquardt, D.W. 1963. An algorithm for the estimation of non-linear parameters. J. SIAM, 11, 431-441.

Mueggler, W.F. 1989. Age distribution and reproduction of intermountain aspen stands. Western Journal of Applied Forestry: 4(2):41-46.

Olson, S.P. and Perala, D.A. 1981. Growth of planted white spruce in relation to soil texture, fertilization, and microenvironment. Final report to the CANUSA-East Spruce Budworm Program.

Payandeh, B. 1974. Formulated site index curves for major timber species in Ontario. Forest Science. 20(2): 143-144.

Rawling, J.O. 1988. Applied regression analysis - a research tool. Wadsworth, Belmont, CA.

Russell, K.W.1963. Plantation white spruce growth related to some forest soil-site characteristics. Plan B paper. St. Paul, MN: University of Minnesota, College of Forestry. 42pp.

SAS Institute Inc. 1992. SAS user's guide: statistics. Version 6 ed., Cary, NC 27512-8000.

Wang, C.H., and Hann, D.W. 1988. Height-diameter equations for sixteen tree species in the central western Willamette valley of Oregon. For. Res. Lab, Oregon State Univ., Res. Pap. 51.

Wykoff, W.R., Crookston, N.L., and Stage, A.R. 1982. User's guide to the stand prognosis model. USDA For. Serv. Gen. Tech. Rep. INT-133.

Chapter 6

Age distribution in regenerated stands

6.1 Introduction

Regenerated stands have become more and more important in Alberta boreal forests, since extensive natural stands have been cut and replaced by regenerated stands. With accelerated use of Alberta's forest resources, there is a need to advance our understanding of dynamics of the regenerated stands. An age distribution model not only exhibits the age structure of newly regenerated stands, but also plays an important role in defining the characteristics of the new stands in a growth simulation model such as MGM (Morton and Titus 1984, Huang 1992). Combining the early height growth model with the age distribution for the new stands allows growth simulation of tree height and age. Simulation of stand development over time can then begin.

White spruce (*Picea glauca* (Moench) Voss), trembling aspen (*Populus tremuloides* Michx), and lodgepole pine (*Pinus contorta* var. *latifolia* Engelm) are three major species in Alberta regenerated stands. For the area in which data were collected, after clear-cutting, the cut-blocks were planted or seeded with different species, or left for natural regeneration. The objective of this study was to model the age distribution for the major species in Alberta regenerated stands during their early years of their development. The data used in this study were from permanent plots on which age was recorded for most individual trees. A two-parameter Weibull distribution was used to model the age distribution for white spruce, aspen, and lodgepole pine. The fit was based on weighted nonlinear least squares method. The fitted models predict the age distributions at 2, 4, 6, 8, 10, and 12 years since regeneration.

6.2 Materials and methods

6.2.1 Data

The data used in this study were provided by the Alberta Land and Forest Services. They are permanent sample plots designed to "monitor" juvenile stand establishment and growth on cut-blocks (Gilday 1990). The data were chosen for this study because age was measured for most individual trees. The plots were established in various ecoregions across the province since 1980. There were 107 plots. They were remeasured every two years and with same plots remeasured up to 7 times. At plot establishment, harvest, regeneration treatment, and other information were recorded for each plot. The regeneration treatments included left-for-natural, seeded, and planted with different species. Of the 107 plots, 69 were planted with white spruce in which 33 were planted with 1-year-old containerized seedlings and 36 were planted with 3-year-old bareroot seedlings; 22 were left for natural regeneration or seeded with lodgepole pine or mixed species; 7 were planted with lodgepole pine; 3 were seeded with white spruce; and 6 were seeded with forage and unspecified species. The last three groups were not of sufficient size for analysis and 6 plots were re-seeded or thinned. They were excluded from this study.

Each plot consisted of 40 quadrats (2 m $\times$ 2.5 m) that were alternatively established on each side of four 25 metre transects distributed over a cut-block (Gilday 1990). White spruce, lodgepole pine, jack pine, black spruce, and all fir species were defined as acceptable species. In a quadrat, each tree of acceptable species and the three tallest deciduous trees were tagged and measured. Measurements included species, age, height, and damage code. Other deciduous trees were counted by species. Tree age was estimated when a tree was first found on the site. The damage code indicated the damage categories for live trees and reported possible causes of mortality when trees were first found dead on the site (Gilday 1990).

Note that, in each 2 m × 2.5 m quadrat, only the first three deciduous trees were tagged and measured in terms of tree age. So the data for aspen were not complete. There may have been more shorter and younger aspen trees than observed in the aspen age distribution.

6.2.2 Tree age distribution summary

For white spruce, the data included plots planted with both 1-year-old containerized spruce and 3-year-old bareroot spruce. The tree ages were different for the two planting methods at planting time, and so at the subsequent years. For example, after 2 years since planting, a plot planted with containerized spruce had 3-year-old planted trees and ingress, but the plot planted with bareroot spruce had 5-year-old planted trees and ingress. Therefore it is not suitable to summarize all the plots based on the planting year. Instead the reference year is the origin year for the oldest planted tree, and age distribution was summarized by the reference years. In this study, the reference year was also called the year since regeneration (YSR), and defined as follows. For white spruce plantations, if a plot was planted with 1-year-old containerized white spruce seedlings, the planting year was designated 1 YSR. If the plot was planted with 3-year-old bareroot seedlings, the planting year was 3 YSR. Since most aspen trees were found in the plots with white spruce, aspen was assigned the same YSR as the spruce. For lodgepole pine, the data included the plots seeded with pure pine or mixed species and the plots left for natural regeneration. They all had scarification preparation before the applied seeding and the expected natural seeding (Gilday 1990). For a lodgepole pine plot, regardless whether it was seeded or left for natural, its scarification treatment year was designated zero year since regeneration (YSR).

Table 6-1 shows the number of plots at progressive YSR for the three species. In odd-numbered years, the number of plots was insufficient for analysis, because the plots were remeasured every two years. For example, only 5, 5, and 6 white spruce plots were found at 5, 7, and 9 YSR, compared with 40, 29, and 14 plots at 6, 8, and 10 YSR. For this reason, the age

distribution was only evaluated at 2, 4, 6, 8, 10, and 12 YSR for white spruce and aspen, and 2, 4, 6, and 8 YSR for lodgepole pine.

At each of 2, 4, 6, 8, 10, and 12 YSR, the total trees per ha and the proportional frequency of trees at each age were summarized (Tables 6-2, 6-3, and 6-4). For example, at 2 YSR, the data included 31 plots containing white spruce. Converted to a per ha basis, there were 673 1-year-old trees, 948 2-year-old trees, 416 3-year-old trees, and so on, i.e., 31% were 1-year-old, 43% were 2-year-old, 19% were 3-year-old, and so on. The summation of the proportional frequencies was equal to 1. The proportional frequencies of trees are also shown in Figures 6-1, 6-2, and 6-3 with the red color columns for white spruce, aspen, and lodgepole pine.

6.2.3 Model specification

The age distributions in Figures 6-1, 6-2, and 6-3 provide a distribution histogram for the three species in the regenerated stands less than 12 YSR. To model these age distribution, the Weibull distribution was fitted to the data. The Weibull distribution has been the most commonly used model to fit distributions of age, diameter, height, basal area, and volume (Clutter *et al.* 1983, Knox and Peet 1989, Burk and Newbarry 1984, Ek *et al.* 1975, Shiver 1988), since Bailey and Dell (1973) demonstrated that diameter distributions in even-aged stands could be quantified by the Weibull distribution. It is preferred due to its flexible form and mathematical tractability (Bailey and Dell 1973).

Let x be age of a tree, and $f(x)$ denote the age probability distribution function. Under the assumption that x is a two-parameter Weibull distribution with scale parameter b and shape parameter c, the probability distribution function has the general form:

$$f(x)=\frac{c}{b}\left(\frac{x}{b}\right)^{c-1}\exp\left[-\left(\frac{x}{b}\right)^{c}\right], \tag{6.1}$$

where $x \geq 0,\ b > 0,\ c > 0,$ and $f(x) = 0,$ elsewhere. The corresponding cumulative distribution function $F(x)$ is:

$$F(x) = \Pr ob(Age < x) = \int_{-\infty}^{x} f(x)dx = 1 - \exp\left[-\left(\frac{x}{b}\right)^{c}\right]. \tag{6.2}$$

The parameter b is close to the mean of the distribution $b\Gamma(1 + 1/c)$, where Γ is the gamma function (Golberg 1984, page 396). The parameter c indicates the skewness of the age distribution (Vanclay 1994). The Weibull approximates a normal distribution for $3.25 \leq c \leq 3.61$, and displays progressively more positive skewness below this range and negative skewness above this range (Knox and Peet 1989).

To fit the Weibull distribution to the data, the parameter b and c need to be estimated. In this study weighted nonlinear regression used for the parameter estimation is described as follows.

6.2.4 Model fitting

Let p_x be the proportional frequency at age x. In this study, age was recorded by year. When an x-year-old tree was found, generally the age was equal to or greater than x years, and less than x+1 years, that is, $x \leq$ Age $< x$+1. Therefore under the Weibull distribution assumption,

$$\begin{aligned} p_x &= \Pr ob(Age < x+1) - \Pr ob(Age < x) \\ &= F(x+1) - F(x) \\ &= \exp\left[-\left(\frac{x}{b}\right)^{c}\right] - \exp\left[-\left(\frac{x+1}{b}\right)^{c}\right]. \end{aligned} \tag{6.3}$$

Based on (6.3), the model to fit was formulated:

$$p_x = \exp\left[-\left(\frac{x}{b}\right)^{c}\right] - \exp\left[-\left(\frac{x+1}{b}\right)^{c}\right] + \varepsilon \tag{6.4}$$

where $\varepsilon \sim N\left(0,\sigma_x{}^2\right)$. Note that the model (6.4) did not have a constant variance, since $\text{var}(p_x) = \dfrac{\hat{p}_x(1-\hat{p}_x)}{n_x}$, where $\hat{p}_x$ was the predicted proportional frequency and n_x was number of trees where $x \leq$ Age $< x+1$ (Wiens 1996). To achieve a homogeneous variance, a weighting factor for fitting model (6.4) was given as $\text{weight} = \dfrac{1}{\text{var}(p_x)} = \dfrac{n_x}{\hat{p}_x(1-\hat{p}_x)}$. Based on the weighting, the sum of squared error (SSE) was equal to

$$SSE = \sum_{x=1}^{n} \frac{n_x\left(\hat{p}_x - \exp\left[-\left(\frac{x}{b}\right)^c\right] - \exp\left[-\left(\frac{x+1}{b}\right)^c\right]\right)^2}{\hat{p}_x(1-\hat{p}_x)} \tag{6.5}$$

where x was the discrete age (x = 1, 2, 3, ….n), and n was the number of discrete ages. The nonlinear least squares estimates of the parameters were obtained by minimizing the sum of squared error (SSE). The procedure PROC NLIN on SAS/STAT (SAS Institute Inc. 1992) was used to accomplish the fitting. The weight was specified $\left(_\text{WEIGHT}_ = \sqrt{n_x/(\text{model}.p_x(1-\text{model}.p_x))}\right)$. The derivative free iterative method was applied (METHOD=DUD), and multiple starting estimates for parameters were provided to ensure that the least squares solution was global solution. The evaluation of the fit was judged by the significance of parameter estimates, the mean squared error (MSE), the coefficient of determination (R^2), and the residual analysis. The R^2 were computed according to the following formula:

$$R^2 = 1 - \frac{\sum_{x=1}^{n}(p_x - \hat{p}_x)^2}{\sum_{x=1}^{n}(p_x - \overline{p}_x)^2} \tag{6.6}$$

where p_x was the actual proportional frequency of trees at age x, $\hat{p}_x$ was the predicted proportional frequency, and $\overline{p}_x$ was the averaged actual proportional frequency.

6.3 Results

The fit statistics are shown in Tables 6-5, 6-6, and 6-7 for white spruce, aspen, and lodgepole pine. They include the estimates of parameters *b* and *c*, the mean squared error (MSE), and the coefficient of determination (R^2). The observed range of tree ages at each YSR is also listed.

With the exception of the age distribution at 12 YSR, MSE was low and R^2 high at each year since regeneration. At 12 YSR, there were only 4 to 5 plots, too few to show the general trend. Estimates of parameters were significant at level $\alpha = 0.05$. The residual plot also showed satisfactory with random pattern of positive and negative residuals and an approximately constant variance. As expected, the values for the scale parameter *b* were close to the mean of the distribution, and values for the shape parameter *c* were quite variable.

6.4 Discussion

The observed and predicted age distributions at each YSR are shown in Figures 6-1, 6-2, and 6-3 for white spruce, aspen, and lodgepole pine. It can be seen that the fitted age distribution follows the trend of the actual distribution.

The three species were regenerated by different regeneration treatments, and show different age distribution dynamics with the passage of time. In the data used for this study, white spruce was reestablished from both planting and natural regeneration, aspen from both root sucking and natural seeding, and all lodgepole pine from natural and applied seeding.

In the white spruce plantations (Figure 6-1), at 2 YSR, most trees are 1 to 3 years old. Some of them were planted, and others were geminated from available seed sources. The age range is relatively narrow because of the prior clear-cutting. With the years passing, the mean of the predicted age distribution increases, and so does the range of age. Juvenile white spruce plantations are usually associated with aspen and other deciduous trees that dominate the site. The reproduction method of root sucking gives aspen an initial advantage over white spruce (Mueggler 1989) by displaying dense regeneration and fast early growth. However, white spruce is a shade tolerant species, and in spite of aspen and possible planted white spruce in upper levels, germination of spruce trees still occurs and these trees survive and grow under the canopy. Therefore, with the mixedwood stand development, age is distributed over a wider range. With the passage of time, the probability distribution function gradually becomes more and more flat and symmetric. At the 10 YSR, the age distribution is close to a Bell-shaped with the mean between 7 and 8. The chart for 12 YSR does not show the general trend because there were only 5 plots.

Figure 6-3 shows the predicted and actual age distribution for lodgepole pine in natural and seeded pine stands. Unlike white spruce, lodgepole pine is distributed over a small range of ages. In natural and seeded pine stands, lodgepole pine relies only on germination of seeds to reestablish. At 2 YSR, most trees are 1 year old. Only a few ingress seedlings appear before the scarification treatment, and most ingress seedlings appear within 2 YSR. As time passes, the mean of the predicted age distribution increases with approximately a constant rate. The mean tree age is approximately 3 at 4 YSR, 5 at 6 YSR, and 7 at 8 YSR. The range of ages remains similar for different YSR. There are a few new ingress seedlings as years pass, but the proportional frequency is very low. The age distribution becomes more symmetric, but not flat. At 8 year since scarification, most trees are between 6 and 8 years old. The lodgepole pine stands show a classic even-aged distribution.

Only up to three deciduous trees were measured in a 2 m × 2.5 m quadrat. The aspen data were, therefore, not complete. On one hand, the aspen age distributions in Figure 6-2 may miss small and young aspen seedlings. On the other hand, since the YSR is defined according to the age of planted spruce and is not determined by either harvesting year or planting year, the summarized and predicted aspen age distribution at each YSR in Figure 6-2 may include too much variation. For example, plots at 5 YSR may include plots planted 2 years ago with bareroot seedlings, and plots planted 4 years ago with container seedlings. If the planting treatments were 2 years after clear-cutting, the summarized plots at 5 YSR covered 4 to 6 years after the clear-cutting. Therefore, the summarized age range of aspen may be broader than the real distributions. With more complete and abundant data, more complete study of aspen age distribution is necessary and possible.

In summary, the fitted age distribution models provide a preliminary view of tree age for white spruce, aspen, and lodgepole pine in Alberta regenerated stands. Due to the data limitation, the predicted age distribution is only related to the years since regeneration (YSR) for white spruce plantations, and seeded and natural pine stands, but not to other characteristics such as seed source availability, harvesting information, other silviculture treatments, site productivity, density, species composition and others which may impact on age distribution significantly. Analyzing the relationships between other characteristics to the age distribution will be possible when more data are available.

Table 6-1. Number of plots at each year since regeneration (YSR) by species

Year since regeneration (YSR)	Number of plots for white spruce	Number of plots for aspen	Number of plots for lodgepole pine
1	24	24	14
2	31	29	15
3	41	30	5
4	51	44	11
5	5	3	4
6	40	34	7
7	5	3	3
8	29	26	7
9	6	5	1
10	14	12	
11	5	5	11
12	5	4	
13	5	3	

Table 6-2. Trees per ha and proportional frequency at each age for white spruce

Years since regeneration	2		4		6		8		10		12	
Tree age	#	%	#	%	#	%	#	%	#	%	#	%
1	673	31	197	8	283	10	190	5	46	1	10	
2	948	43	318	13	328	12	274	8	68	2	170	7
3	416	19	343	14	279	10	505	14	225	6	200	8
4	44	2	1047	43	304	11	457	13	357	9	110	4
5	60	3	430	18	260	9	274	8	604	15	200	8
6	24	1	56	2	755	27	334	9	611	15	170	7
7	13	1	16	1	428	15	233	6	318	8	120	5
8	6		17	1	38	1	662	18	536	13	60	2
9	3		4		104	4	476	13	171	4	80	3
10	3		4		19	1	40	1	625	16	110	4
11	2		2		5		88	2	196	5	210	8
12	5		3		4		22	1	50	1		
13			3		1		5		164	4	620	25
14			5		1		3		32	1		
15					4		2		4		460	18
16			1		4		2		4			
17			2				5					
18					1		5					
19					3				7			
20	2						2		7		10	

Table 6-3. Trees per ha and proportional frequency at each age for aspen

Years since regeneration	2		4		6		8		10		12	
Tree age	#	%	#	%	#	%	#	%	#	%	#	%
1	609	18	673	20	175	5	113	4	38	1	313	7
2	1491	43	836	25	97	3	63	2	263	9	613	13
3	864	25	650	19	574	16	227	8	154	5	513	11
4	400	12	795	24	965	26	79	3	33	1	650	14
5	43	1	257	8	659	18	431	15	263	9	125	3
6	21	1	111	3	728	20	810	27	42	1	113	2
7			14		310	9	567	19	704	25	288	6
8			3		122	3	381	13	796	28	25	1
9			1		16		233	8	350	12	100	2
10			1				48	2	88	3	1063	23
11							13		83	3	475	10
12									8		288	6
13											75	2

Table 6-4. Trees per ha and proportional frequency at each age for lodgepole pine

Years since regeneration	2		4		6		8	
Tree age	#	%	#	%	#	%	#	%
1	6120	75	332	3	157	2	50	1
2	1807	22	550	6	236	2	129	1
3	253	3	6473	67	671	7	243	3
4	30		1923	20	829	9	286	3
5			327	3	6293	66	643	7
6			36		1200	13	779	9
7					136	1	5371	61
8					14		1164	13
9							129	
10							14	

Table 6-5. Fit statistics for white spruce age distribution models

Year since regeneration	Number of plots	Estimate of parameter *b*	Estimate of parameter *c*	MSE	R^2	Actual range of age
2	31	2.599169	3.084225	0.000098	0.9955	[1, 12]
4	51	4.738014	5.503727	0.001713	0.8790	[1, 17]
6	40	6.387831	2.566869	0.002149	0.6434	[1, 19]
8	29	7.282156	2.094930	0.001157	0.6814	[1, 21]
10	14	8.458395	2.706834	0.000957	0.7259	[1, 22]
12	5	14.408008	10.637479	0.002278	0.5510	[1, 22]

Table 6-6. Fit statistics for aspen age distribution models

Year since regeneration	Number of plots	Estimate of parameter *b*	Estimate of parameter *c*	MSE	R^2	Actual range of age
2	29	3.010080	3.231350	0.001136	0.9625	[1, 7]
4	44	3.700673	2.222110	0.001089	0.9174	[1, 10]
6	34	5.697416	3.519104	0.001118	0.8830	[1, 10]
8	26	7.383390	4.579053	0.001263	0.9842	[1, 11]
10	12	8.517325	7.064689	0.003365	0.6555	[1, 12]
12	4	8.604191	1.387640	0.004804	0.0813	[1, 13]

Table 6-7. Fit statistics for lodgepole pine age distribution models

Year since regeneration	Number of plots	Estimate of parameter *b*	Estimate of parameter *c*	MSE	R^2	Actual range of age
2	15	1.858056	5.635233	0.000450	0.9975	[1, 4]
4	11	3.836540	9.788841	0.000815	0.9901	[1, 6]
6	7	5.743717	13.476371	0.001520	0.9733	[1, 8]
8	7	7.742991	16.680875	0.001707	0.9554	[1, 10]

Figure 6-1. Actual and predicted proportional frequency of age for white spruce
(red - actual, green - predicted)

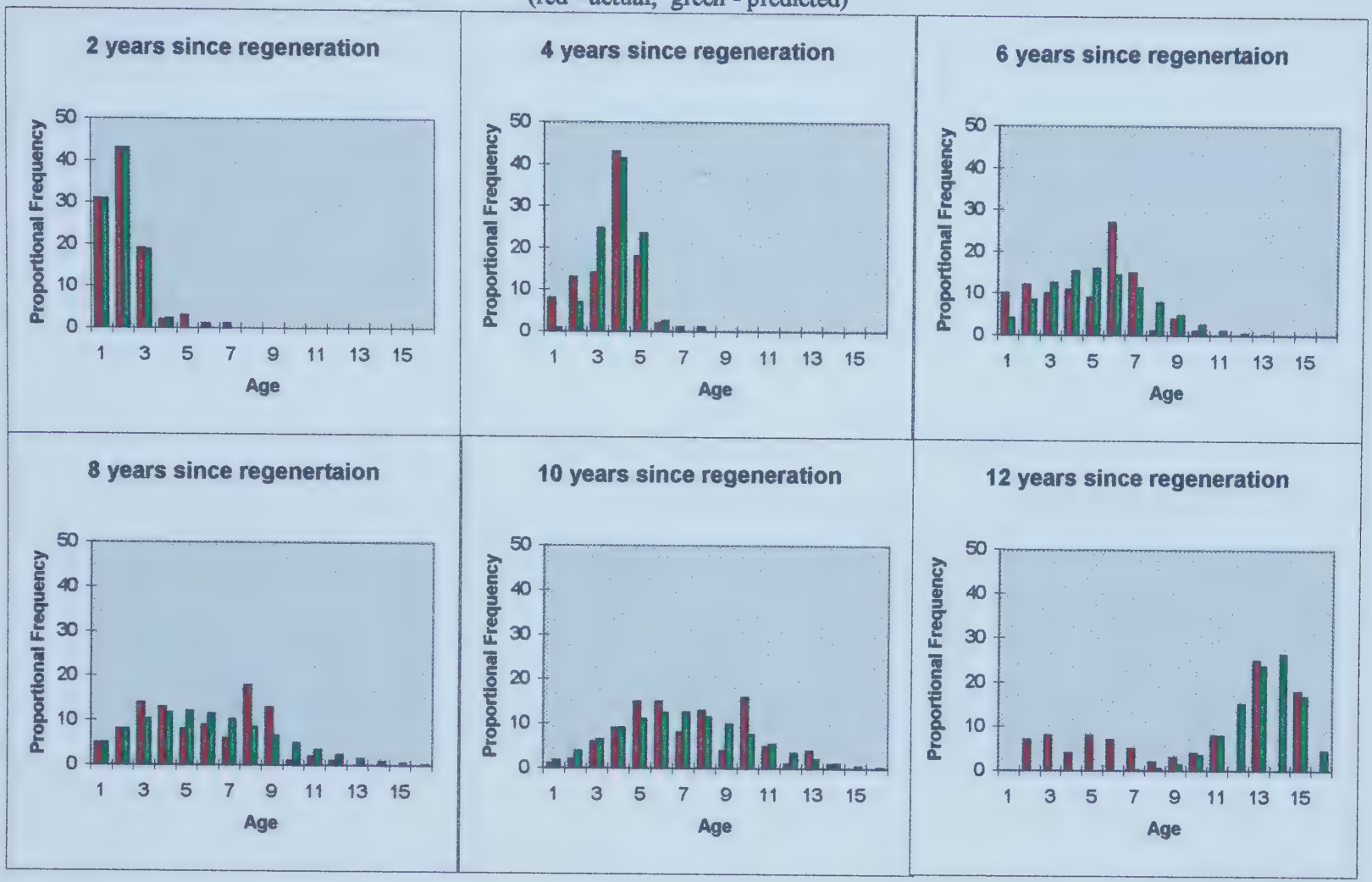


Figure 6-2. Actual and predicted proportional frequency of age for aspen
(red - actual, green - predicted)

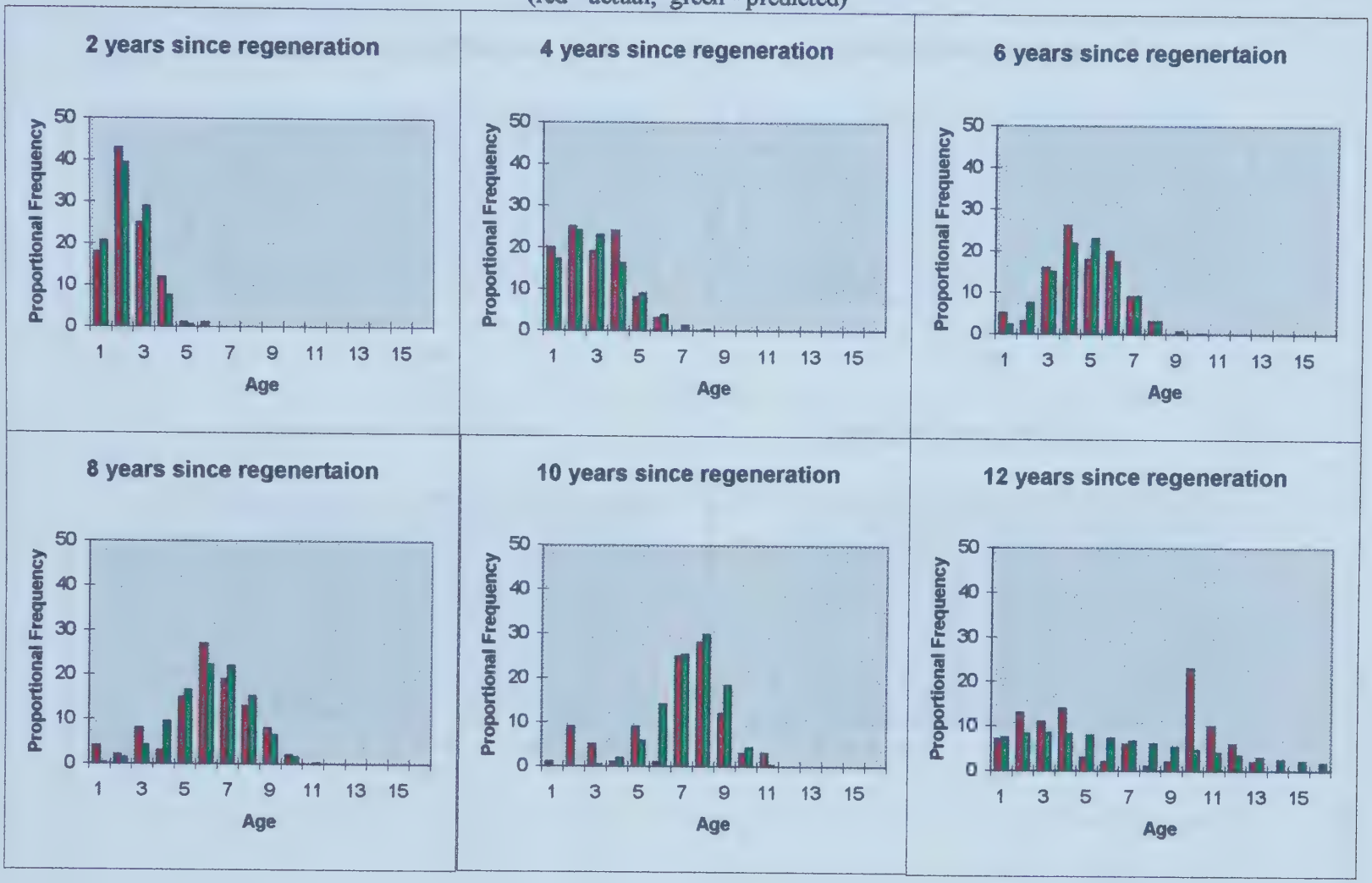
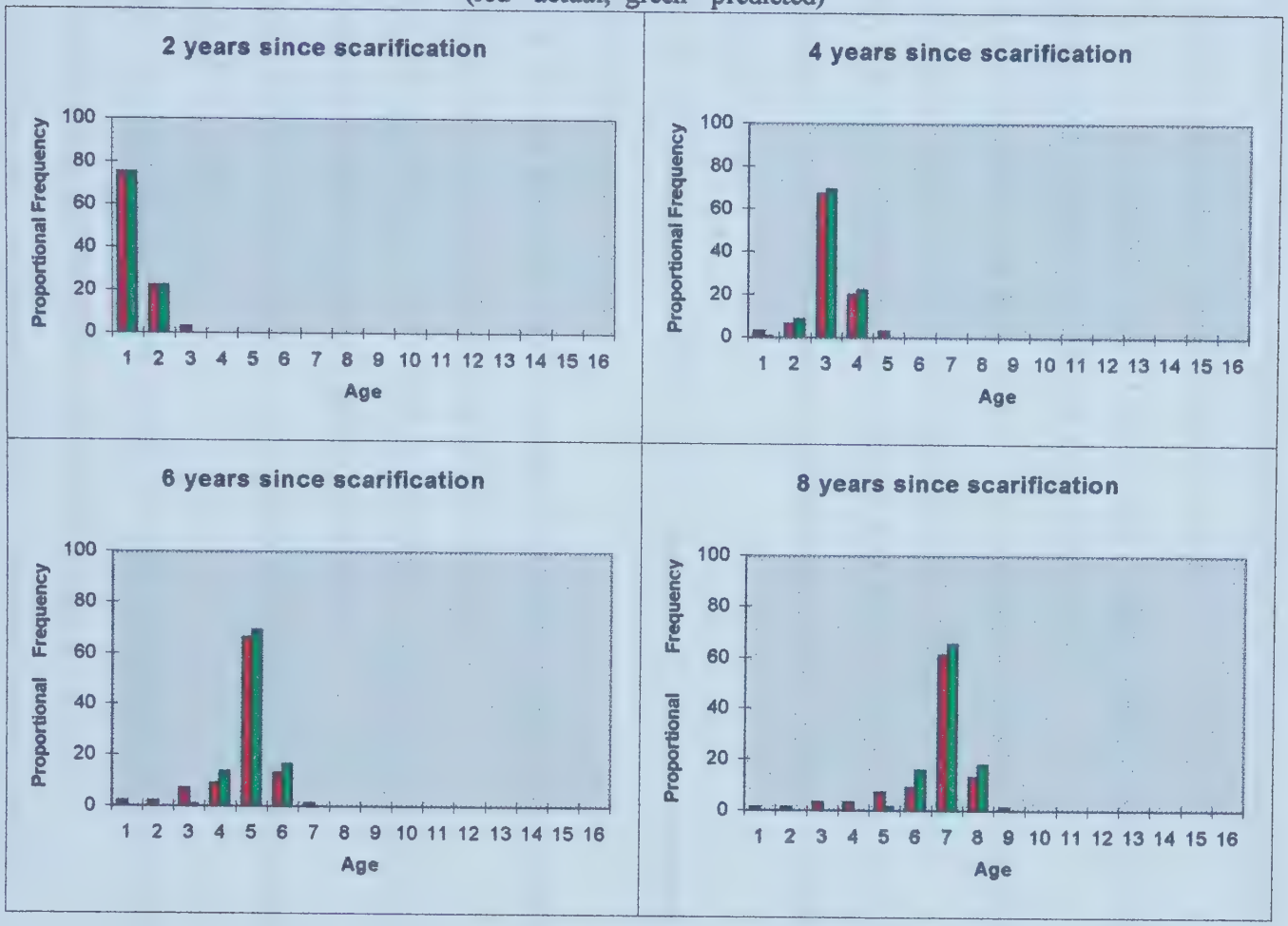


Figure 6-3. Actual and predicted proportional frequency of age for lodgepole pine (red - actual, green - predicted)



6.5 References

Bailey, R.L. and T.R. Dell. 1973. Quantifying diameter distributions with the Weibull function. Forest Science. 19:97-104.

Burk, T.E. and J.D. Newbarry. 1984. A simple algorithm for moment-based recovery of Weibull distribution parameters. Forest Science, 30: 329-332.

Clutter, J.L., J.C. fortson, L.V. Pienaar, G.H. Brister, and R.L.Bailey. 1983. Timber management: a quantitative approach. John Wiley & sons.

Ek, A.R., J.N. Issos, and R.L. Bailey. 1975. Solving for Weibull diameter distribution parameters to obtain specified mean diameters. Forest Science, 21:290-292.

Gilday, D.W. 1990. Monitor Plot survey manual. Published by Alberta Land and Forest Service.

Golberg, M.A. 1984. An introduction to probability theory with statistical applications. Plenum press, New York.

Huang, S. 1992. Diameter and height growth models. Ph.D. Thesis, Department of Forest Science, University of Alberta.

Knox, R. G., and Peet, R.K. 1989. Population dynamics in loblolly pine stands: changes in skewness and size inequality. Ecology, 70(4):1153-1166.

Morton, R.T. and Titus, S.J. 1984. The development of a mixedwood stand growth model. Final report to Forest Research Branch, Alberta Forest Service.

Mueggler, W.F. 1989. Age distribution and reproduction of intermountain aspen stands. Western J. Appl. For., 4(2):41-46.

Vanclay. J.K. 1994. Modeling forest growth and yield - Applications to mixed tropical forests. CAB international, Wallingford, Oxon OX10 8DE, UK.

SAS Institute Inc. 1992. SAS user's guide: statistics. Version 6, first ed., Cary, NC.

Shiver, B.D. 1988. Sample sizes and estimation methods for the Weibull distribution for unthinned slash pine plantation diameter distributions. Forest Science, 34: 809-814.

Wiens, P.D. 1996. Personal communication. Department of Mathematical Science, University of Alberta.

Chapter 7

Individual tree diameter increment models

7.1 Introduction

The boreal forests in Alberta typically have a mixed-species composition. Most growth models suitable for mixedwood stands predict individual tree diameter growth based on tree and stand variables (Vanclay 1994). These individual tree diameter increment models are essential for individual tree based growth and yield systems in mixedwood stands such as FOREST (Ek and Monerud 1974), PROGNOSIS (Stage 1973, Wykoff 1990), and NORM (Vanclay 1994), and MGM (Morton and Titus 1984, Huang 1992).

Generally a diameter increment model is fitted to data from permanent sample plots on which all trees have been individually identified and remeasured. Several approaches to develop diameter increment models are found in literature. The model can be derived by taking the first derivative of the cumulative diameter prediction function (Daniels and Burkhart 1975), or approached by predicting the stand increment and distributing this among the stems comprising the stand (Clutter and Allison 1974, Leary *et al.* 1979). Another approach models diameter increment by predicting potential growth and using a modifier function to estimate actual increment (Ek and Monserud 1974, Leary *et al.* 1979, Amateis *et al.* 1989). The potential growth is estimated for the trees without competition, and the modifier, the competitive adjustment factor, reduces the potential growth to the predicted level. This approach may be robust, but in irregular mixedwood stands, the dominant age-site relations are difficult to apply and potential growth cannot be reliably estimated (Wykoff 1990). In this study a fourth approach was used. A function is developed that relates diameter increment to tree and stand characteristics including tree size, its relative competitiveness, stand competition, and site productivity (Wykoff 1990,

Martin and Ek 1984, Vanclay 1991). This function can be used in a simulation model to portray stand development over time.

In some studies, basal area increment has been modeled instead of diameter increment (Wykoff 1990), as diameter increment and basal area increment are related mathematically, their modeling methodologies are essentially the same (West 1980, Shifley 1987).

Huang and Titus (1995) developed an individual tree diameter increment model based on the Box-Lucas function for white spruce stands older than 30 years in Alberta. That model was unable to adequately predict diameter growth for juvenile stands (Huang and Titus 1995). It consistently underestimated the diameter growth for small trees. This study is an extension of that work to include stands less than 30 years. Three species were evaluated: white spruce (*Picea glauca* (Moench) Voss), trembling aspen (*Populus tremuloides* Michx), and lodgepole pine (*Pinus contorta* var. *latifolia* Engelm) growing in Alberta mixedwood stands. The data for this study were from permanent sample plots in both mature and juvenile stands. The model predicts individual tree diameter increment based on tree diameter at breast height, total height, total basal area per ha for larger trees, stand basal area per ha, species composition, and site productivity.

7.2 Data

Three data sets from permanent sample plots were used in this study. They are described here using their project names: Permanent Sample Plots (PSP), Stand Dynamics (SDS), and Regenerated Permanent Sample Plots (RPSP). Both PSP and SDS were provided by the Alberta Land and Forest Services, and RPSP were from Grande Prairie Operations, Weyerhaeuser Canada Ltd.

The PSP plots were established and remeasured since the early sixties, and most of them were initially 30 years or older. The SDS and RPSP plots were designed to provide long term

data regarding the growth and dynamics of regenerated juvenile stands. The SDS plots were established in 1983, and the RPSP plots in 1989.

7.2.1 Permanent Sample Plots (PSP)

Over 600 permanent sample groups have been established and remeasured. PSP groups were randomly located throughout the inventory areas of the province to provide representative information for a variety of stand density, species composition, and site condition. A description of the establishment and data collection can be found in Permanent sample plots: the field procedure manual (Alberta Forest Service 1990). Each PSP group consisted of 1 or 4 plots. Each plot has been remeasured up to 5 times over different time intervals. At each measurement time, trees were measured with respect to species, diameter at breast height, and condition code. Approximately 30 percent of trees were measured for height.

7.2.2 Stand Dynamics (SDS)

SDS plots were randomly located under different ecological zones of the province to provide representative information for a variety of silviculture, site classification and forest management needs. There were 207 plots, covering from 1 to 8 years old stands at the first measurement. Each plot was remeasured every two years and up to 5 times. A description of the plot layout and data collection can be found in Stand dynamics system field remeasurement manual (Alberta Forest Service 1993). Each Stand Dynamics plot was a nested plot and included a tree plot, a sapling plot, and a regeneration plot. The tree plot was a 0.1 ha square. Once a stem reached 9.1 cm in diameter at breast height, it was tagged and remeasured for diameter and height. The sapling plot, a square, located in the northwest quarter of the tree plot, occupied 0.025 ha. Trees within the sapling plot greater than 1.1 cm diameter at breast height were tagged and remeasured for diameter and height. There were 4 circular regeneration plots located in the sapling plot. Each was 0.001 ha. All

seedlings within the regeneration plots were recorded excluding current year growth (germinates). Up to 50 seedlings were remeasured for height. If there were more than 50 seedlings, a height class table by species was used to measure the tree heights and count the trees (Appendix 5).

7.2.3 Regenerated Permanent Sample Plots (RPSP)

The plot layout and data collection of RPSP data was quite similar to SDS plots and is described in Permanent sample plot field procedure manual (Weyerhaeuser Canada Ltd. 1993). The plots were located throughout the management area near Grande Prairie, Alberta on a predetermined systematic grid basis. There were 72 plots, and each plot has been remeasured every two years, and up to 3 times. A concentric nested plot design was also employed which placed four circular (0.001 ha) regeneration plots inside a square (0.02 ha) sapling plot that, in turn, was inside a square (0.08 ha) tree plot. In a regeneration plot, all trees which were less than 1.3 m in height were tagged and measured for height. In a sapling plot, trees, not less than 1.3 m in height and not greater than 5 cm in diameter, were tagged and measured for diameter and height. In a tree plot, trees greater than 5 cm in diameter were tagged and remeasured for diameter and height.

For analysis, the three data sets were merged together. Only the first plot of each PSP group was used due to the excessive size of the data set. Each non-overlapping growth period from the remeasurements defined a growth interval. A total number of 1652 growth intervals were obtained from the merged data for white spruce, 8533 for aspen, and 6330 for lodgepole pine. The annual diameter increment was calculated as the difference between the diameter at the end of the growth interval and beginning of the interval divided by the length of the interval. The length of growth interval was computed as: $L = Year_2 - Year_1 + (Adj_2 - Adj_1)$, where $Year_1$ and $Year_2$ are consecutive measurement years, and Adj_1 and Adj_2 are the consecutive month

adjustments. The month adjustment was defined according to the biological growth period in the boreal mixedwood stands (Huang 1992). It was equal to 0.0, 0.2, 0.5, 0.9, and 1 when month was ≤ 4, $=5$, $=6$, $=7$, ≥ 8, respectively.

The annual diameter increment was plotted against diameter at the beginning of the growth interval. An example for white spruce is shown in Figure 7-1. The scatter plot showed an extremely large variation for the diameter increment. In this case, selecting an appropriate base curve of the model mostly relied on approximating the general biological process of diameter growth.

Repeated measurements on the same trees tended to be correlated, and the data did not meet the assumption of independence. Since the number of the multiple measurements for the same trees was very small in comparison with the number of trees, the problem was not expected to be serious (Vanclay 1994). The data were used to fit the model, and the serial correlation problem was considered after the model fitting.

7.3 Model Specification

Selecting an appropriate base function is the first step to model diameter increment directly. The conceptual framework for the increment model is the positively skewed, unimodal diameter increment curve that is typical of tree growth processes (Wykoff 1990) and compatible with the sigmoidal S-shaped yield curves. Diameter increment increases to a maximum early in the life of a tree and then decreases slowly, approaching zero as the tree matures. To model this relationship, an increment function used in Wykoff (1990) and Vanclay (1991) was considered:

$$DIN = \alpha DBH^{\beta} \exp\left(-\gamma DBH^{2}\right), \tag{7.1}$$

where *DIN* is diameter increment, *DBH* is diameter at breast height, and α, β, and γ are unknown parameters. This curve begins at value zero, increases with increasing *DBH*, reaches the

maximum value at $\sqrt{\beta/2\gamma}$, and then decreases and approaches 0 as $DBH \rightarrow +\infty$. In equation (7.1), with increasing *DBH*, simultaneously DBH^{β} increases and $\exp(-\gamma DBH^2)$ decreases. When *DBH* is less than $\sqrt{\beta/2\gamma}$, the increasing DBH^{β} term dominates the curve. After that, the decreasing $\exp(-\gamma DBH^2)$ becomes more and more dominant, and drives the curve asymptotically towards zero. The parameter α does not impact on the turning point, but it influences the slope of the curve, and affects the magnitude of the maximum. Several typical graphs of equation (7.1) are shown in Figure 7-2.

The curve has zero intercept at *DBH* zero. However when a tree is 1.3 meter tall (*DBH*=0), its biomass is not zero, so the model with zero intercept may underestimate increment for small trees (Vanlay 1994). Adding an intercept term in model (7.1) was attempted in a preliminary study, however no significant difference was found near *DBH* zero. The reason might be that in model (7.1), when $0 < \beta < 1$, the general range of β in the diameter increment model, $\lim_{DBH \rightarrow 0} \frac{dDIN}{dDBH} = +\infty$. This shows that *DIN* can increase dramatically once *DBH* passes zero, and the model can portray a rapidly increasing increment for trees with *DBH* near zero.

A similar evaluation shows the difficulty in applying the Box-Lucas function (Huang and Titus 1995) to juvenile trees. The Box-Lucas function is:

$$DIN = \frac{\varphi}{\varphi - \psi}\left[\exp(-\psi DBH) - \exp(-\varphi DBH)\right]. \qquad (7.2)$$

where φ and ψ are positive parameters. Like equation (7.1), it is also a positively skewed, unimodal curve, and has zero intercept at *DBH* zero. This model was used successfully in modeling individual tree diameter increment for larger trees in Alberta mixedwood stands (Huang and Titus 1995), but in application it underestimated the diameter increment for juvenile trees. When *DBH*=0, the first derivative of *DIN* in the Box-Lucas equation is equal to the value

of φ and is generally less than 1. This property limits the acceleration of the increment curve at *DBH* zero. Other variables may be able to be incorporated into φ to increase diameter increment for small trees. However, it is likely that the plain Box-Lucas function will underestimate the increment for small trees. This is the main reason that the equation (7.1) was chosen as the base curve instead of the Box-Lucas function (7.2), although the later function performed well for the large trees (Huang and Titus 1995).

In equation (7.1), the diameter increment increases as either α or β increases, and decreases as γ increases. This property makes it easy to modify and interpret the model according to the biological growth pattern.

Another advantage of equation (7.1) is that the statistical properties of this model are close-to-linear (Ratkowsky 1990). With a log transformation, the model becomes:

$$\log(DIN) = \log(\alpha) + \beta \log(DBH) - \gamma DBH^2. \tag{7.3}$$

The linearization brings the non-linear function (7.1) within the bounds of simple regression methodology. This linear model belongs to the family of equations described by Vanclay (1994) as

$$\log(DIN) = \log(\alpha) + \beta \log(DBH) - \gamma DBH^k$$

where typically $k = 1$ or $k = 2$. Zeide (1993) examined many growth equations and recommended the equation with $k = 1$ as diameter increment model. Vanclay (1991) also used the equation with $k = 1$ to model diameter increment for tropical rain forests. Wykoff (1990) used this base curve with $k = 2$ to predict basal area increment in the PROGNOSIS model. A preliminary analysis in this study showed that the function with $k = 1$ performed well within the data range, but it predicted excessive increment for very large trees. Therefore the equation with $k = 2$ was chosen. The DBH^2 term hastens the approach to zero for large *DBH* without destroying the essential advantage of the model (Wykoff 1990).

A shortcoming of the linearized model (7.3) is that, to fit the model, the observations with zero and negative diameter increment have to be excluded due to the logarithm transformation. However not all the negative diameter increments are caused by measurement error. Zero increment and small decrement may be caused by natural variation in stem size and by bark shedding (Vaclay 1994). Failure to accommodate these observations may lead to overestimates of diameter increments and timber yield. Vanclay (1991) used $\log(DIN + 0.02)$ instead of $\log(DIN)$ to avoid the deletion.

In this study, the linear model (7.3) was fitted without zero and negative increment to provide the initial guesses of the parameters α, β and γ, and the nonlinear technique was used to fit the nonlinear model (7.1) to all the data. Other tree and stand level characteristics were incorporated into α, β and γ by using the parameter prediction method (Clutter *et al.* 1983, Huang and Titus 1995). The final model was formulated as

$$
\begin{aligned}
& DIN = \alpha DBH^{\beta} \exp\left(-\gamma DBH^{2}\right) \\
& \alpha = f(\text{other variables}) \\
& \beta = g(\text{other variables}) \\
& \gamma = h(\text{other variables})
\end{aligned}
\qquad (7.4)
$$

and the fitting was accomplished using the procedure PROC REG and the procedure PROC NLIN on SAS/STAT software (SAS Institute Inc. 1992).

7.4 Explanatory variables

7.4.1 Diameter at breast height and total height

Diameter at breast height in cm (*DBH*) was chosen as the primary explanatory variable, since it is almost universally available and relates to the maturity of the tree. Maturity is closely

related to the increment. Total tree height (*HEIGHT*) was also considered for predicting diameter increment.

7.4.2 Stand basal area

As a measure of stand competition, total basal area (m^2) per ha (*BA*) was used, since it includes effects of both size and density of trees in a stand.

7.4.3 Total basal area in larger trees

For an individual tree, to present its competitive stresses, total basal area of trees larger than the subject tree (*GGR*) was considered in this study. Computed as

$$(7.5) \qquad GGR = 0.0001 \sum_{DBH_i > DBH} \frac{\pi}{4} DBH_i^2 ,$$

where DBH_i was the *i*-th diameter greater than the diameter of the subject tree (*DBH*). *GGR* is nonnegative, and presents the position of the subject tree in the tree basal area distribution. If the tree has the largest diameter (or basal area), GGR is zero. Trees with smaller diameter have greater *GGR*. Wykoff (1990) and Vanclay (1991) found *GGR* or the ratio of *GGR* to total basal area a useful predictor of diameter increment.

7.4.4 Species composition

BA and *GGR* assume that all species exert equal competitive influence, but in mixedwood stands some species could exert a greater influence than others. Species composition (SC_{SP}) was considered as an additional explanatory variable indicative of mixedwood competition. It was defined as follows:

$$(7.6) \qquad SC_{SP} = \frac{BA_{SP}}{BA},$$

where BA_{SP} was the total basal area per ha for the species under consideration, and *BA* was the total basal area per ha for all species.

7.4.5 Site

Site index based on stand age and height was not used in this study because age was difficult to measure accurately and was not available for most trees. Instead, a species specific site productivity index (*SPI*), defined as average height of the dominant and codominant trees at 20 cm reference diameter (Huang and Titus 1993), was used as a site productivity measure in this study. The 20 cm reference diameter corresponds roughly to the 50 year reference age in traditional Alberta site index curves (Huang and Titus 1993).

Since the only historical site record was the Phase 3 Inventory for the juvenile stands (SDS and RPSP), it was used to identify the site productivity classification for the cut-blocks. Almost all the cut-blocks located in good and medium sites. If the site class was good, *SPI* was designated 22. If it was medium, then *SPI* was equal to 18.

7.5 Results

7.5.1 Model fitting

First, the linearized model (7.3) was fitted excluding observations with zero and negative diameter increment. The linear least squares fit of equation (7.3) was accomplished using the PROC REG procedure on SAS software (SAS Institute Inc. 1992) (Appendix 6). *HEIGHT*, *BA*, *GGR*, SC_{SP}, and *SPI* were also incorporated into the intercept term $\log(\alpha)$ of the linear model (7.3). Since using stepwise regression analysis to select variables often included an unnecessarily large number of variables, and the fitted model performed poorly when it was used near the limits of the data (Vanclay 1994), the maximum R^2 improvement technique (MAXR) developed by Goodnight (1979, as referenced in SAS Institute Inc. 1992) was utilized in variable selection

(SELECTION=MAXR). The MAXR method did not select a single model. Instead, it tried to give the best one-variable model, the best two-variable model, and so forth (SAS Institute Inc. 1992). Checking all the models, the one with most reasonable statistical and biological properties was selected.

The selected linear model for white spruce was

$$\log(DIN) = \left[\log(\alpha_1) + \alpha_2 BA + \alpha_3 SPI + \alpha_4 GGR + \alpha_5 SC_{SW} + \alpha_6 \log(BA)\right] + \beta \log(DBH) - \gamma DBH^2, \tag{7.7}$$

where SC_{SW} was white spruce composition, and α_1, α_2, α_3, α_4, α_5, α_6, β, and γ were parameters. Comparing equation (7.7) to (7.4), the final relationship between parameter α and other variables was

$$\alpha = \alpha_1 \exp\left[\alpha_2 BA + \alpha_3 SPI + \alpha_4 GGR + \alpha_5 SC_{SW} + \alpha_6 \log(BA)\right]. \tag{7.8}$$

With the initial estimates for the parameters from the linear fit, the nonlinear regression was used to fit the model

$$\begin{aligned} DIN &= \alpha DBH^{\beta} \exp\left(-\gamma DBH^2\right) \\ \alpha &= \alpha_1 \exp\left[\alpha_2 BA + \alpha_3 SPI + \alpha_4 GGR + \alpha_5 SC_{SW} + \alpha_6 \log(BA)\right] \\ \beta &= \beta_1 \\ \gamma &= \gamma_1 \end{aligned} \tag{7.9}$$

to the entire data set. Furthermore, *HEIGHT*, *BA*, *GGR*, SC_{SP}, and *SPI* were also tried to form the parameters β and γ of equation (7.9) for any further improvement of the fit. It was found that incorporating *HEIGHT* into parameter γ gave a better fit. The final diameter increment model for white spruce was:

$$\begin{aligned} DIN &= \alpha DBH^{\beta} \exp\left(-\gamma DBH^2\right) \\ \alpha &= \alpha_1 \exp\left[\alpha_2 BA + \alpha_3 SPI + \alpha_4 GGR + \alpha_5 SC_{SW}\right] BA^{\alpha_6} \\ \beta &= \beta_1 \\ \gamma &= \gamma_1 + \gamma_2 / HEIGHT. \end{aligned} \tag{7.10}$$

A positive sign of α_1, α_2, α_3, α_4, α_5, α_6, or β_1 indicated a positive effect on diameter increment, and a positive sign of γ_1, and γ_2 indicated a negative effect on diameter increment.

This procedure was similar to the method of parameter prediction used for a Box-Lucas function (Huang and Titus 1995) and Weibull type function (Clutter *et al.* 1983), in which the parameters of the function were related to tree and stand characteristics, but the form of the base function remained unchanged. With the similar analysis, the final diameter increment model for aspen was found:

$$
\begin{aligned}
& DIN = \alpha DBH^{\beta} \exp\left(-\gamma DBH^2\right) \\
& \alpha = \alpha_1 \exp\left[\alpha_2 SPI + \alpha_3 GGR + \alpha_4 SC_{AW}\right] BA^{\alpha_5} \\
& \beta = \beta_1 \\
& \gamma = \gamma_1
\end{aligned}
\qquad (7.11)
$$

where SC_{AW} was aspen composition. The final model for lodgepole pine was:

$$
\begin{aligned}
& DIN = \alpha DBH^{\beta} \exp\left(-\gamma DBH^2\right) \\
& \alpha = \alpha_1 \exp\left[\alpha_2 GGR + \alpha_3 SC_{PL}\right] BA^{\alpha_4} SPI^{\alpha_5} \\
& \beta = \beta_1 \\
& \gamma = \gamma_1 + \gamma_2 / HEIGHT
\end{aligned}
\qquad (7.12)
$$

where SC_{PL} was lodgepole pine composition.

The preliminary fits of equations (7.10), (7.11), and (7.12) were accomplished using the PROC NLIN procedure on SAS/STAT software (SAS Institute Inc. 1992) (Appendix 6). The Marquardt iterative method (Marquardt 1963) was applied to solve the nonlinear least squares equations.

7.5.2 Residual examination

Plotting the studentized residuals against predicted diameter increment, it was found that the variance of the residuals decreased with increasing predicted value. To achieve homogenous error variance, the weighted nonlinear least squares technique was applied. Several functions of

DBH, such as the exponential, logarithm, and power functions, were used as the weighting factor, and the plot of studentized residuals against the predicted diameter increment was examined in each case. Empirically it was found that the model with weight $w_i = DBH^{0.2}$ produced estimates with satisfactory residual plots. The plot for white spruce is shown in Figure 7-3. It displays an approximately homogenous band of the error variance.

7.5.3 Serial correlation

Since the data used to fit the models included multiple remeasurements of the same individuals, the residuals might contain serial correlation. To check the correlation, the data were sorted by time order for each individual. The empirical one-step serial correlation was examined by plotting each residual, except the first one, against the one preceding it for all individuals (Draper and Smith 1981). The one-step serial plot exhibited no significant tendency of the trend. To view the correlation for more steps, similar plots for residuals two steps apart, three steps apart, and so forth, were made (Appendix 6). No significant tendency of the trend was found. In an empirical study of yield models, Borders *et al.* (1988) also found no serial correlation when the data derived from non-overlapping growth interval.

The one-step serial correlation can also be detected by the Durbin-Watson test (Draper and Smith 1981). However simply selecting Durbin-Watson test option in PROC NLIN is not appropriate in this study. The Durbin-Watson test in PROC NLIN is under the assumption that the data is a time series, that is, the whole data presents one individual tree measured at consequent time points. But the data used in this study were merged by three data sets including thousands of trees. It contained many time series instead of one.

7.5.4 Final fits

The procedure PROC NLIN on SAS/STAT software (SAS Institute Inc. 1992) was used to fit equations (7.10), (7.11), and (7.12) for white spruce, aspen, and lodgepole pine with the weighting factor $DBH^{0.2}$. The final fit statistics including the nonlinear least squares estimates for the parameters, the asymptotic standard errors and asymptotic 95% confident intervals of the estimates, the mean squared errors (MSE), and the coefficient of determination (R^2) are listed in Tables 7-2, 7-3, and 7-4. The R^2 was calculated as

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(DIN_i - \hat{DIN}_i\right)^2}{\sum_{i=1}^{n} \left(DIN_i - \overline{DIN}_i\right)^2} \tag{7.13}$$

where DIN_i was the observed increment and $\hat{DIN}_i$ was the predicted increment for the i-th tree (i=1, 2, ..., n), and $\overline{DIN}$ was the averaged observe increment.

All the parameter estimates were significant at $\alpha = 0.05$ level. The signs of parameter α and β were positive with any input values, and the sign of parameter γ was positive once the tree height was large enough, all of which ensured that the diameter increment predictions relative to diameter had the desired unimodal, and positively skewed shape. According to the R^2, more than 40 percent of the weighted diameter increment variation was explained by the fitted model for white spruce and lodgepole pine, and about 15 percent for aspen.

7.6 Discussion

The fitted diameter increment models as expressed by equations (7.10), (7.11), and (7.12) provide individual tree diameter increment predictions for white spruce, aspen, and lodgepole pine in Alberta juvenile and mature mixedwood stands, given tree diameter at breast height, total height, total basal area per ha for all species, total basal area per ha of larger trees,

species composition, and site productivity index. The model also confirms some commonly held beliefs and growth trend for the three species.

As expected, the effect of stand basal area (*BA*), as presented by parameter α_2 and α_6 in equation (7.10) for spruce, α_5 in equation (7.11) for aspen, and α_4 in equation (7.12) for lodgepole pine, is negative. The stand basal area measures stand competition. With more stand competition, less diameter increment is predicted for the individuals.

The individual competitiveness was presented by the total basal of trees larger than the subject tree (*GGR*). The corresponding parameters, α_4 in equation (7.10) for white spruce, α_3 in equation (7.11) for aspen, and α_2 in equation (7.12) for lodgepole pine, are all negative. It shows that, with other factors constant, larger trees in the stand, as measured by smaller *GGR*, exhibit larger diameter increments.

With more stand basal area, and more trees larger than the individual being considered, less diameter increment is expected for the individual. This also indicates that increments are greatest for dominant trees in low density stands and least for suppressed trees in high density stands.

In aspen and spruce mixedwood stands, the shade intolerant aspen shows faster early growth and rapidly forms dominance on the site by occupying the upper layer of the canopy while white spruce in the understory exhibits slow juvenile growth. For about 50 years, aspen continues to dominate the site, but white spruce becomes progressively more conspicuous in the understorey. As time passes, the competitiveness of aspen is reduced relative to that of white spruce as individual aspen trees start to die at 60 to 80 years of age, giving dominance gradually to the more shade tolerant and longer-lived white spruce (Peterson and Peterson 1992). This dynamic process results in decreasing *GGR* for white spruce. The diameter increment model shows that white spruce, with decreasing *GGR*, exhibits fast growth after release from aspen

dominance in the upper layer. *GGR* has also been used to show the thinning response to diameter increment in Wykoff (1990).

Species composition has a negative coefficient in equation (7.10) for white spruce (α_5), in equation (7.11) for aspen (α_4), and in equation (7.12) for lodgepole pine (α_3). This indicates that the species composition negatively impacts diameter increment, i.e., when other factors are the same, trees have larger diameter increment in mixedwood stands than in pure stands. Pure stands are likely to be even-aged associated with high density. More competition in high density may limit the diameter growth.

For the two conifer species, white spruce and lodgepole pine, the significant positive coefficients of the site productivity index, α_3 in equation (7.10) and α_5 in equation (7.12), indicate that the site productivity has a positive impact on the diameter increment. Similar relationships were found for red pine plantations (Martin and Ek 1984) and tropical rain forests (Vanclay 1991). A better site supports a faster diameter growth. An opposite result was obtained from the fitted aspen model by the negative α_2 in equation (7.11). Site has a negative impact on the aspen diameter increment. The reason is unclear, but a possible explanation may be that the fast growing aspen on good sites results in more competition in the stands. Aspen is a shade intolerant species. More competition may encourage height growth, but limit the diameter growth.

In summary, an individual tree diameter increment model for both juvenile and mature trees was developed. The data were from both juvenile stands and mature stands. Using an appropriately selected unimodal-shaped diameter increment base function, and the method of parameter prediction (Clutter *et al.* 1983, Huang and Titus 1995), the diameter increment model was elaborated as a function of tree diameter, height, stand total basal area, total basal area of

larger trees, species composition, and site productivity. The selected base function prevented the model from underestimating the diameter increment for small trees, and the nonlinear regression method allowed the data including trees with zero increment and small decrement that might have been caused by natural variation in stem size.

Table 7-1 Summary statistics for both mature and juvenile trees

Species	N	Variable	Mean	Std Dev	Minimum	Maximum
White spruce	1652	*DIN*	0.295266	0.215140	0.010000	1.450000
		DBH	20.963075	13.853424	0.200000	63.300000
		HEIGHT	16.755841	9.716467	1.310000	37.900000
		BA	29.765947	16.065840	0.000800	62.213500
		GGR	15.874371	12.777291	0.000000	56.399000
		SC_{SW}	0.530611	0.286819	0.001200	1.000000
		SC_{AW}	0.289297	0.260828	0.000000	0.990000
		SPI	17.201573	2.109483	10.400000	25.100000
Aspen	8533	*DIN*	0.342503	0.207984	0.010000	1.500000
		DBH	3.180159	6.267563	0.100000	66.900000
		HEIGHT	3.871432	4.841195	1.170000	32.900000
		BA	3.723846	8.389610	0.000500	59.137800
		GGR	2.129968	4.930577	0.000000	45.301300
		SC_{SW}	0.018573	0.078843	0.000000	0.970000
		SC_{AW}	0.841672	0.216971	0.002200	1.000000
		SPI	18.154611	1.098298	11.500000	25.100000
Pine	6330	*DIN*	0.482251	0.305552	0.010000	1.500000
		DBH	4.664706	7.194419	0.200000	47.600000
		HEIGHT	4.246552	5.792654	1.200000	31.400000
		BA	9.102467	12.240107	0.003900	62.213500
		GGR	4.380706	6.258236	0.000000	52.960500
		SC_{PL}	0.803736	0.251079	0.004100	1.000000
		SPI	18.026050	0.892238	10.700000	22.000000

Table 7-2. The parameter estimates for white spruce diameter increment model (7.10)

Parameter	Estimate	Asymptotic Std. Error	Asymptotic 95% Confidential Interval Lower	Upper	MSE	R^2
α_1	0.248199	0.033187	0.183113	0.313284	0.027533	0.4212
α_2	-0.028495	0.002349	-0.033103	-0.023888		
α_3	0.035541	0.006753	0.022296	0.048786		
α_4	-0.003306	0.002290	-0.007797	-0.001186		
α_5	-0.374701	0.050448	-0.473638	-0.275763		
α_6	-0.114308	0.0125569	-0.138934	-0.089682		
β_1	0.337148	0.0271902	0.283824	0.390473		
γ_1	0.000461	0.000086	0.000291	0.000630		
γ_2	-0.008395	0.002600	-0.013495	-0.003295		

Table 7-3. The parameter estimates for aspen diameter increment model (7.11)

Parameter	Estimate	Asymptotic Std. Error	Asymptotic 95% Confidential Interval Lower	Asymptotic 95% Confidential Interval Upper	MSE	R^2
α_1	0.575212	0.087975	0.402781	0.747643	0.0534	0.1461
α_2	-0.037530	0.007834	-0.052880	-0.022181		
α_3	-0.035116	0.002608	-0.040232	-0.030005		
α_4	-0.143009	0.008357	-0.159395	-0.126630		
α_5	-0.130973	0.007026	-0.144743	-0.117204		
β_1	0.374670	0.012878	0.349429	0.399911		
γ_1	0.000569	0.000046	0.000479	0.000659		

Table 7-4. The parameter estimates for lodgepole pine diameter increment model (7.12)

Parameter	Estimate	Asymptotic Std. Error	Asymptotic 95% Confidential Interval Lower	Asymptotic 95% Confidential Interval Upper	MSE	R^2
α_1	0.116423	0.048722	0.020899	0.211925	0.068660	0.4326
α_2	-0.085100	0.004378	-0.093682	-0.076518		
α_3	-0.234787	0.022462	-0.278820	-0.190753		
α_4	-0.100062	0.006860	-0.113510	-0.866142		
α_5	0.598386	0.143437	0.317200	0.879571		
β_1	0.296014	0.021733	0.253411	0.338618		
γ_1	0.003181	0.000186	0.0028154	0.003546		
γ_2	-0.016407	0.004027	-0.024300	-0.008513		

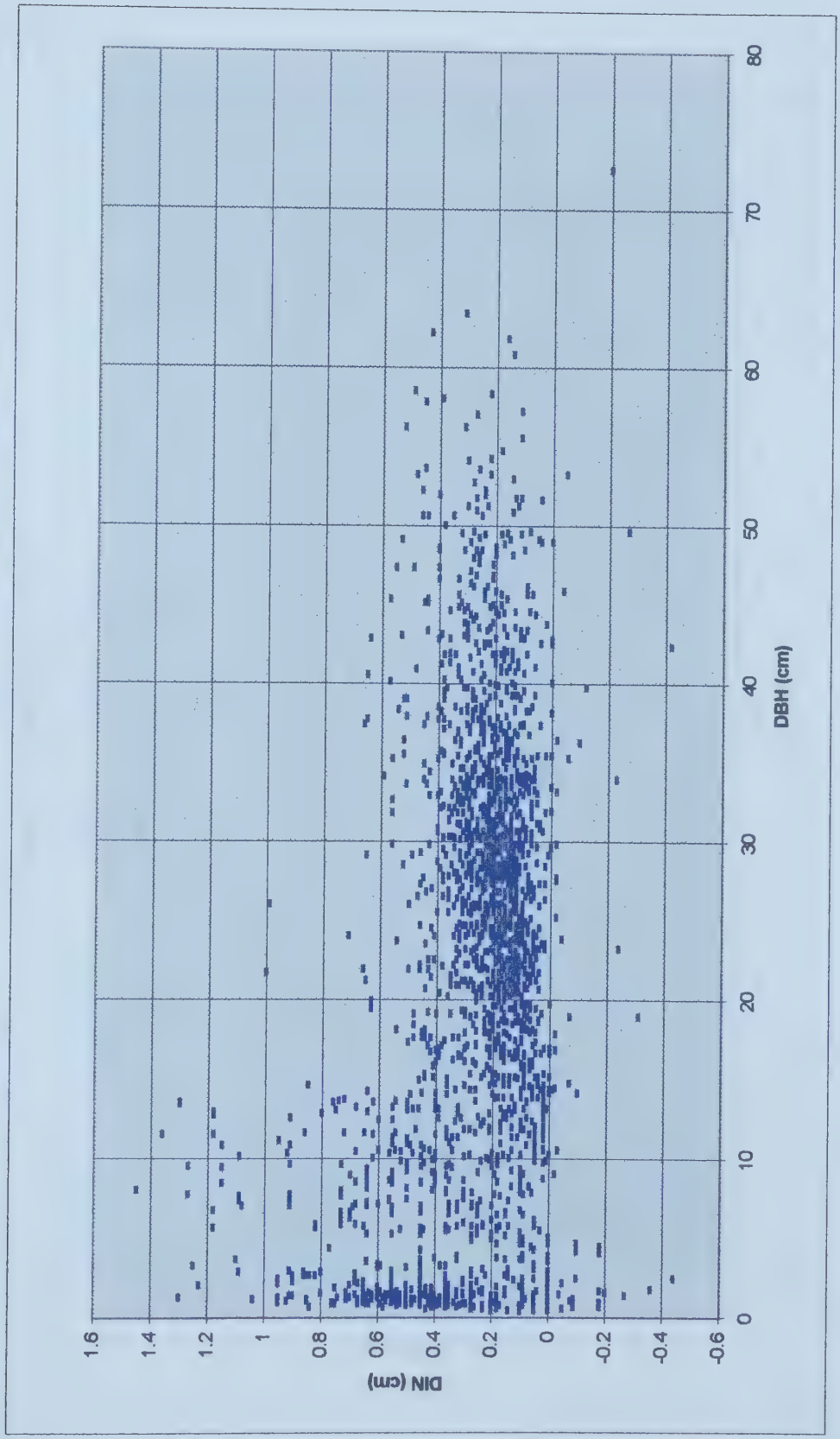


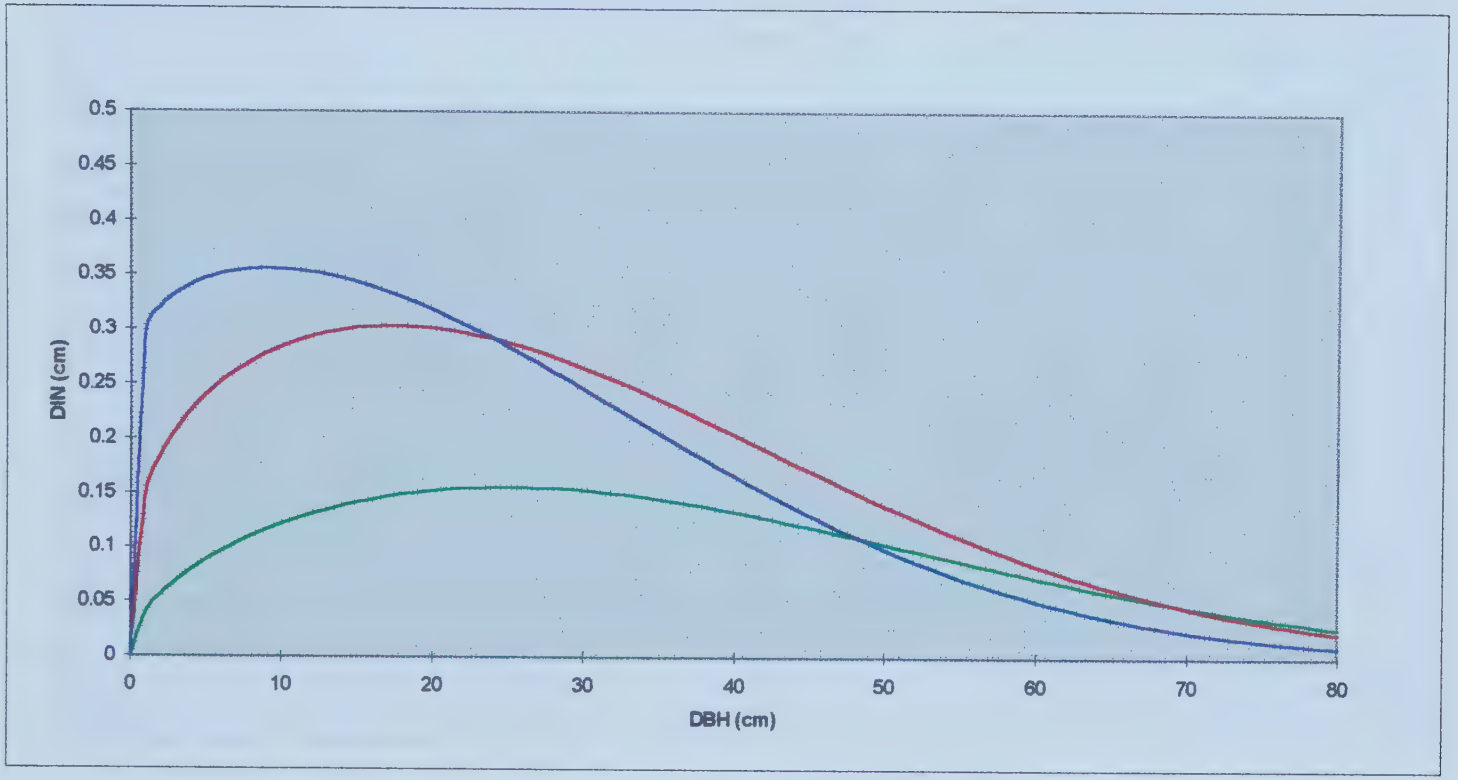
Figure 7-1. Plot of diameter increment against diameter for white spruce trees

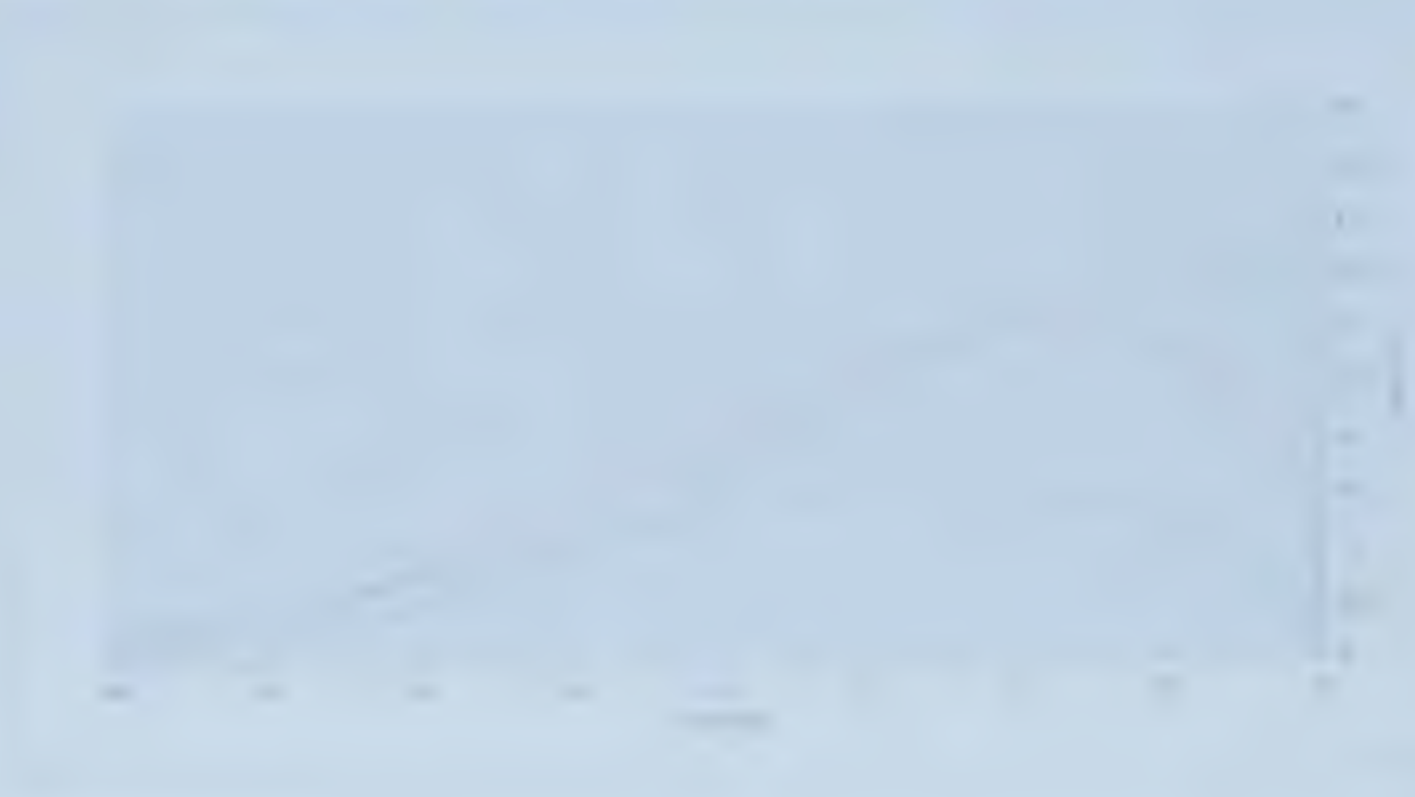
Figure 7-2. Typical graphs of equation (7.1) by varying parameters α and β

green: $\alpha = 0.04$ and $\beta = 0.5$,

red: $\alpha = 0.15$ and $\beta = 0.3$,

blue: $\alpha = 0.3$ and $\beta = 0.3$





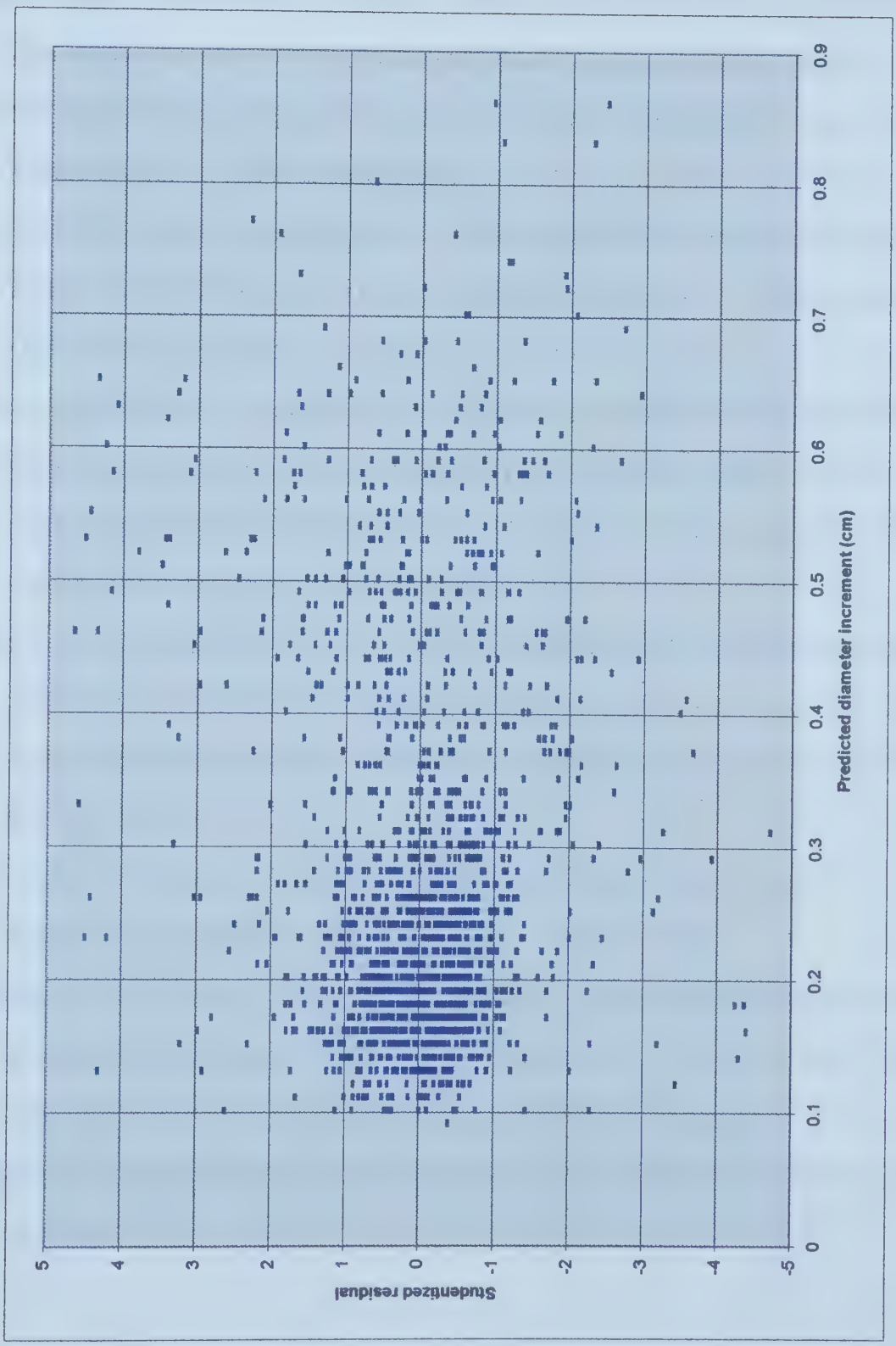


Figure 7-3. Plot of studentized residual against predicted diameter increment for white spruce model (7.10)

7.8 References

Alberta Forest Service. 1990. Permanent sample plots: field procedure manual. Timber Management Branch, Alberta Forest Service, Edmonton, Alberta. FMOPC 83-03.

Alberta Forest Service 1993. Stand dynamics system field remeasurement manual. Timber Management Branch, Alberta Forest Service.

Amateis, R.L., Burkhart, H.E. and Walsh, T.A. 1989. Diameter increment and survival equations for loblolly pine trees growing in thinned and unthinned plantations on cutover, site-prepared lands. Southern J. Appl. For., 13(4): 170-174.

Border, B.E., Bailey, R.L. and Clutter, M.L., 1988. Forest growth models: Parameter estimation using real growth series. In A.R. Ek, S.R. Shifley and T.E. Burk (ed.) forest Growth Modeling and Prediction. Proc. IUFRO Conf., 23-27 Aug. 1987, Minneapolis, MN. USDA For. Serv., Gen. Tech. Rep. NC-120, pp. 660-667.

Clutter, J.L. and Allison, B.J., 1974. A growth and yield model for Pinus radiata in New Zealand. In J. Fries (ed.) Growth Models for Tree and Stand Simulation. Proc. IUFRO Working Party S4.01-4 meetings, 1973. Dep. For. Yield Res., Royal Coll. For., Stockholm. Res. Note 30. 379p.

Clutter, J.L., J.C. fortson, L.V. Pienaar, G.H. Brister, and R.L.Bailey. 1983. Timber management: a quantitative approach. John wiley & sons. New York.

Daniels, R.F., and Burkhart, H.E. 1975. Simulation of individual tree growth and stand development in managed loblolly pine plantations. Div. For. Wild. Res., Virginia Polytechnic Inst. And State Univ., Blacksburg, VA. FWS-5-75. 69 pp.

Draper N.R. and Smith, H. 1981. Applied regression analysis. 2nd edition. John Wiley and Sons, New York.

Ek, A.R. and Monserud, R.A., 1974. Trial with program FOREST: growth and reproduction simulation for mixed species even- or uneven-aged forest stands. In J. Fries (ed.) Growth Models for Tree and Stand Simulation. Proc. IUFRO Working Party S4.01-4 meetings, 1973. Dep. For. Yield Res., Royal Coll. For., Stockholm. Res. Note 30, pp. 56-69.

Goodnight, J.H. 1979. A tutorial on the sweep operator. American statistician, 33:149-158.

Huang, S. 1992. Diameter and height growth models. Ph.D. Thesis, Department of Forest Science, University of Alberta.

Huang, S, and Titus, S.J. 1993. An index of site productivity for uneven-aged and mixed-species stands. Can. J. For. Res., 23: 558-562.

Huang, S. and Titus, S.J. 1995. An individual tree diameter increment model for white spruce in Alberta. Can. J. For. Res., 25: 1455-1465.

Leary, R.A., Holdway, M.R. and Hahn, J.T., 1979. Diameter growth allocation rule. In: A generalized forest growth projection system applied to the lake states region. USDA For Serv., Gen. Tech. Rep. NC-49, pp. 39-46.

Martin, G.L. and Ek, A.R. 1984. A comparison of competition measures and growth models for predicting plantation red pine diameter and height growth. Forest Science, 30(3): 731-743.

Marquardt, D.W. 1963. An algorithm for the estimation of non-linear parameters. J. SIAM, 11, 431-441.

Morton, R.T. and Titus, S.J. 1984. The development of a mixedwood stand growth model. Final report to Forest Research Branch, Alberta Forest Service.

Peterson E.B. and Peterson, N.M. 1992. Ecology, management, and use of aspen and balsam poplar in the prairie provinces. Forestry Canada.

Ratkowsky, D.A. 1990. Handbook of nonlinear regression. Marcel Dekker, Inc., New York.

SAS Institute Inc. 1992. SAS user's guide: statistics. Version 6, first ed., Cary, NC.

Stage, A.R. 1973. Prognosis model for stand development. USDA For. Serv. Res. Pap. INT-137. 32 p.

Shifley, S.R. 1987. A generalized system of models forecasting central states tree growth. USDA For. Res. Pap. NC-279. 10 p.

Vanclay, J.K. 1991. Aggregating tree species to develop diameter increment equations for tropical rainforests. Forest ecology and management, 42: 143-168.

Vanclay. J.K. 1994. Modeling forest growth and yield - Applications to mixed tropical forests. CAB international, Wallingford, Oxon OX10 8DE, UK.

West, P.W. 1980. Use of diameter increment and basal area increment in tree growth studies. Can. J. For. Res. 10: 71-77.

Weyerhaeuser Canada Ltd. 1993. Permanent sample plot field procedure manual. Unpublished report. Grande Prairie Operations, Weyerhaeuser Canada Ltd., Alberta.

Wykoff, W.R. 1990. A basal area increment model for individual conifers in the northern rocky mountains. Forest Science, 36(4): 1077-1104.

Zeide, B. 1993. Analysis of growth equations. Forest Science, 39(3): 594-616.

Chapter 8

General discussion and conclusion

Under the framework of the individual tree distance-independent growth and yield model, this study has accomplished the three primary objectives stated in Chapter 1. First, a generalized logistic model of individual tree mortality was developed for mature mixedwood stands. The effects of tree diameter, diameter increment, stand basal area, species composition, and site productivity on mortality were evaluated, and the biological interpretation of the model was discussed. Secondly, based on the permanent sample plot data collected in regenerated stands, the basic growth relationships for the major tree species were analyzed, which included: the individual tree mortality model, the individual tree height versus age relationship, and the age distribution model at each year since regeneration. The impacts of regeneration method, damage agents, stand density, species composition, and site productivity on juvenile tree growth and mortality were evaluated. Third, diameter increment was also evaluated for both juvenile and mature trees, and a prediction model was developed. The above relationships made it possible to outline a juvenile stand growth model, which can provide a preliminary description of the juvenile stand dynamics.

The main contributions of this study are:

1. Development of the generalized logistic model of individual tree mortality for Alberta mixedwood stands. The generalized logistic model, treating survival as a compound interest phenomenon, was shown to be appropriate to data with unequal remeasurement intervals. The maximum likelihood estimation procedure was derived in this study, and implemented using the quasi-Newton numerical method. The fitted model related the survival probability

to tree diameter, diameter increment, stand basal area, species composition, and site productivity, described the natural phenomenon of mortality in mixedwood stands, and supported a logical biological interpretation of relationships between the survival probability and the explanatory variables.

2. Application of the lifetime distribution model on tree survival data. The methodology of the lifetime distribution analysis was systematically reviewed and applied to the juvenile tree mortality. A nonparametric analysis was used to portray the survival data, and provided a reasonable theoretical model of the lifetime distribution. A parametric distribution model was developed relating the lifetime distribution to tree age, height, damage agents, regeneration method, stand density, and site productivity. No such comprehensive analysis and discussion of the lifetime distribution model has been found in previous tree mortality studies. Based on the comparison of the logistic model with the parametric lifetime distribution model, it is recommended that if the mortality prediction mainly focuses on the periodic survival probability, as commonly used in a stand growth model system, the logistic model is more applicable in terms of the easy model construction, and the tractable and direct model interpretation and application. If the survival pattern and the risk of mortality need to be viewed as time passes, and the survival probability needs to be predicted for a longer term, the lifetime distribution model is preferred.
3. Based on the permanent sample plot data collected from the regenerated stands younger than 15 years, the following growth and mortality relationships were developed:
    - The height versus age relationships were evaluated for the major species in Alberta juvenile stands. A truncated exponential function was used as a base function to model juvenile tree growth until 1.3 metre. The parameter prediction method was used to incorporate additional tree and stand level variables such as regeneration method, damage agent, stand density, species composition, and site productivity into

the parameters of the based function. The model is unique in that it includes variables such as regeneration method, damage agents, and species composition, to an extent that most of the previous models have not.

- Age distribution at years since regeneration was summarized and modeled with a two-parameter Weibull distribution. Weighted nonlinear regression was used to estimate the parameters of the Weibull distribution by minimizing the weighted sum of squared errors between actual and predicted proportional frequencies of trees at ages. The fitting method has not been found in the literature for modeling tree age and size distributions. It shows a satisfactory fit to the data and it is easy to apply due to the mathematical tractability of the Weibull distribution.
- The logistic model of juvenile tree mortality was developed for the regenerated stands. The model predicts the survival probability in two years given tree age, height, regeneration method, damage agents, stand density, species composition, and site productivity.

4. Development of periodic diameter increment models for the major species in both regenerated and mature stands. Using an appropriately selected base function, and the method of parameter prediction, the diameter increment model relates individual tree diameter increment to tree diameter, total height, stand basal area, total basal area of larger trees, species composition, and site productivity. The selected base function prevented the model from underestimating the diameter increment for small trees. The nonlinear regression method allowed the data including trees with zero increment and decrement that might have been caused by natural variation in stem size, where previous studies, using the linear regression to the corresponding linearized model, excluded those observations and risked an overestimation of the diameter increment.

5. Based on the fitted relationships, a juvenile stand growth model can be outlined as follows. At a given year since regeneration, with given density and species composition, the age distribution model allows simulation of a tree list with age for each tree. The tree factor, the number of trees per ha represented by a single tree in the tree list, can be calculated from the stand density and the species composition. The juvenile height versus age relationship predicts height for each individual on the initial tree list, and allows the trees to grow up to 1.3 metre. The mortality model predicts the survival probability, and allows an adjustment of the tree factor for each tree. Once a tree reaches 1.3 m breast height, the diameter increment model drives the tree to grow.

The models developed in this study also confirm some commonly held theories and beliefs about mixedwood stand dynamics.

1. Mortality in juvenile stands younger than 15 years

   At establishment, small natural or planted seedlings are vulnerable to many damage agents, and are susceptible to death, thus mortality in this early stage is relatively high. Suppression was found to be a common cause of mortality, although the suppression may not be so much from other trees as from grass and shrubs. Other causes of mortality were also discovered such as climatic extremes, flooding, insects, and *etc*. As trees grew taller, their ability to resist competition and damage agents improved, and the survival probability increased. Better growth on richer sites also assisted seedlings to survive damage agents, so better site productivity improved early juvenile tree survival. Stand density, reflecting competition among trees, had a negative impact on tree survival. Lodgepole pine survived better in pure stands than in mixedwood stands. In aspen and white spruce mixedwood stands, after clear-cutting, reproduction by root sucking resulted in aspen achieving high density and fast early growth, leading to domination of the site by occupying the upper layer

of the canopy. Aspen had a competitive advantage over the shade tolerant spruce which survived with slow juvenile growth. However, excessive aspen density may reduce white spruce survival. For white spruce, natural seedlings had slight higher survival probability than container planted seedlings. Compared with container and natural seedlings, the bareroot seedlings had lower survival probability especially when they were young and small.

2. Mortality in natural stands older than 30 years

In the stands older than 30 years, small trees were less competitive and more likely to be suppressed, thus having lower survival probability. With increasing diameter, the survival probability increased to a point, then stayed relatively stable, and then decreased for very large diameter. Old trees with very large diameters may decline in vigor, and their survival probability decreased as a result of reduced ability to survive damage. Trees, of any size, with greater diameter increment had higher survival. It was found that trees had lower survival probability on richer sites, which was the opposite to that in juvenile stands. The reason might be that the better establishment and fast early growth on richer sites lead to earlier competition among the trees. In turn the increased stand basal area resulted in greater competition and reduced tree survival. As in juvenile stands, lodgepole pine had higher survival in pure stands than in mixedwood stands. In aspen and white spruce mixedwood stands older than 30 years, aspen continued to dominate the site, but white spruce became progressively more conspicuous in the understorey. White spruce had a high and stable survival probability, likely reflecting its shade tolerant nature and ability to withstand competition from aspen. When the aspen reached about 30 metre height, mortality probability increased as the tree became more and more sensitive to the effects of factors like diseases, insects, wind, and physical damage. The longer-lived white spruce eventually dominated as the short-lived aspen died out.

3. Height in juvenile stands younger than 15 years

In juvenile stands, seedlings showed an accelerated initial growth if the site was appropriate and the seedlings were free of competition and damage agents. As is generally acknowledged, trees grow faster on richer sites and slower in very dense stands. Aspen achieved faster early growth than white spruce or lodgepole pine. Container planted white spruce was taller than bareroot planted white spruce of the same age. Both container and bareroot seedlings grew faster than natural regeneration. At early ages, lodgepole pine was taller than natural spruce and shorter than planted spruce, but later on, the height of lodgepole pine exceeded the planted spruce and showed a high potential increase. In aspen and white spruce mixedwood juvenile stands, although excessive aspen density may lead to growth stagnation for juvenile white spruce, suitable aspen density, providing appropriate shading of young white spruce seedlings, may help protect the young white spruce seedlings from damage agents and have a positive effect on their height growth.

4. Recruitment

In natural and seeded pine stands, most lodgepole pine trees appeared within 2 years after scarification. Over time the range of ages remained narrow indicating little or no on-going recruitment. The lodgepole pine stands showed a classic even-aged distribution. Juvenile white spruce plantations were usually associated with aspen and other deciduous trees that dominated the site. However, white spruce is a shade tolerant species, and, in spite of aspen and other deciduous trees in upper levels, the germination of spruce trees still occurred and these ingress trees survived and grew under the canopy. The uneven age distribution of juvenile stands indicated that recruitment of spruce continued for some time.

5. Diameter increment

Individual tree diameter increment increased rapidly at early ages, reached a maximum, and then decreased as the tree matured. Generally diameter increment was larger

on richer sites. With more stand basal area, less diameter increment was expected. Trees that were large relative to others in a stand exhibited larger diameter increment. Therefore, increments were greatest for dominant trees in low density stands and least for suppressed trees in high density stands. Overall, trees had larger diameter increment in mixedwood stands than in pure stands. The reason might be that pure stands are more likely to be even-aged associated with high density which may limit diameter growth. In aspen and white spruce mixedwood stands, after release from aspen dominance in the upper layer, white spruce showed increased diameter growth. When white spruce started to replace aspen in canopy level, aspen, exhibited decreased diameter increment.

Discussion and recommendations for further study:

The juvenile height versus age relationships, the age distribution model, and the logistic model of juvenile individual tree mortality are all based on the permanent sample plots younger than 15 years. With additional data, the three models need to be extended to older regenerated stands. The Chapman-Richards function may be preferred over the truncated exponential function when more data are available. Analyzing the effect of tree and stand characteristics on age distribution will be possible with more comprehensive data. The logistic model of mortality may be retained, unless the remeasurement interval becomes different and the generalized logistic model will be used.

The site productivity classification of the parent stands was used for the juvenile stands in this study. However, after harvesting, the site may change. Therefore, the juvenile stand site productivity needs to be analyzed. Ingrowth is an important component in stand dynamics, especially in succession. It is the part that the current MGM model is missing, and needs to be studied.

With more data available for regenerated stands, comparisons of the growth characteristics between natural stands and regenerated stands will be possible. Any suggestions and conclusions obtained from the comparison will be helpful to interpret the change of the forest resources, and guide regenerated stand management and silviculture activities.

There are numerous possibilities for extensions and improvements to the present study, especially with more data collected from the regenerated stands. Nevertheless, the objectives of this study were met, and the models should be beneficial to the understanding of the mortality process in natural mixedwood stands and the dynamics of the regenerated stands, and lead to better management of Alberta forest resources.

Appendices

Appendix 1. Splus code for fitting the generalized logistic model based on the maximum likelihood estimation

The following code provides an example for fitting the generalized logistic model (2.2) based on the maximum likelihood estimation. The data file "swsplus.dat", summarized from the original permanent sample data (PSP), contains individual white spruce trees, on each of which, the value of the explanatory variable *DBH*, *DIN*, *BA*, *SPI*, SC_{AW}, *L*, and the binary response variable *Y* are listed. In this program, *beta0* must be given as initial estimates of the parameters, and *f* is defined as the negative log likelihood function. The call to the nonlinear minimizer, *nlmin*, minimizes the function *f*, and then obtains the maximum likelihood estimates for the generalized logistic model. The output results are listed in *mle*.

```
#################################################################################################
#Input data and abstract variables from the columns;
#DBH = diameter at breast height,
#DIN = annual diameter increment,
#BA = stand basal area per ha,
#SPI = site productivity index,
#SC_AW = aspen species composition,
#L  = length of the remeasurement interval,
#Y = life condition code at the end of the interval.
        data <- matrix (scan ("swsplus.dat"), ncol=13 , byrow=T)
        Y <- data[,13]
        L <- data[,12]
        DBH <- data[,7]
        BA <- data[,9]
        SPI <- data[,10]
        SC_AW <- data[,11]
        DIN <- data[,8]
#Calcualte DBH^2 and DBH^2/BA;
        DSQUARE <-DBH*DBH
        INT <-DBH*DBH/BA
#Form the matrix X;
        x<-cbind(1, DBH, DSQUARE, DIN, BA, SPI, INT)
#Set the initial values for parameter vector β;
        beta0<-c(5.7, 0.0677, -0.0010, 0.5724, -0.00756, -0.1056, -0.0101)
#Define the minus log likelihood function according to (2.10);
        f <- function(beta)
```

Appendix 2. Splus code for Hosmer-Lemeshow goodness-of-fit test

The following code illustrates the Hosmer-Lemeshow goodness-of-fit test for the fitted generalized logistic model (2.2). The data file "swsplus.dat", summarized from the original permanent sample data (PSP), contains individual white spruce trees, on each of which, the value of the explanatory variable *DBH*, *DIN*, *BA*, *SPI*, SC_{AW}, *L*, and the binary response variable *Y* are listed. *Beta* is set the maximum likelihood estimates of the generalized logistic model, and *p* is the predicted survival probability for individual trees. The data is sorted by *p* and divided into *g* group. As in the procedure PROC LOGISTC on SAS/STAT, ten groups is used here. The Hosmer-Lemeshow statistics (*hs*) and the p-value (*pvalue*) are calculated.

```
##############################################################################

#Input data and abstract variables from the columns;
#DBH = diameter at breast height,
#DIN = annual diameter increment,
#BA = stand basal area per ha,
#SPI = site productivity index,
#SCAW = aspen species composition,
#L  = length of the remeasurement interval,
#Y = life condition code at the end of the interval.
        data <- matrix (scan ("swsplus.dat"), ncol=13 , byrow=T)
        y <- data[,13]
        l <- data[,12]
        DBH <- data[,7]
        BA <- data[,9]
        SPI <- data[,10]
        SCAW <- data[,11]
        DIN <- data[,8]
#Calcualte DBH^2 and DBH^2/BA;
        DSQUARE <-DBH*DBH
        INT <-DBH*DBH/BA
#Form the matrix X;
        x<-cbind(1, DBH, DSQUARE, DIN, BA, SPI, INT)
#Set the maximum likelihood estimate to parameter vector β;
        beta<-c(4.6045447559, 0.0444855897, -0.0006322475, 7.0645942283, 0.0116165774, -
        0.1072888448, -0.014489933)
#Calculate the predicted survival probability;
        p <- (1/(1+exp(-x%*%beta)))^l
 #Sort data by the predicted survival probability;
        yp <- cbind(y,p)
        ypsort <- yp[sort.list(yp[,2]),]
#Assign the number of groups for the Homser and Lemeshow test;
#Generally 10 groups was divided as in PROC LOGISTIC on SAS/STAT 1992;
```

```
        g<-10
#Calculate the number of all observations;
        n <- length(y)
#Calculate the number of observations in each group except the last one;
        m <- trunc(n/g+0.5)
#Calculate the Homser and Lemeshow statistic according to (2.14);
#Assign zero to the coming variables;
        pred<-rep(0,g)
        avgp<-rep(0,g)
        act<-rep(0,g)
        chi<-rep(0,g)
# Calculate the actual and predicted number of trees for group 1 to g-1;
#Calculate the average predicted survival probability for group 1 to g-1;
#Calculate the Homser and Lemeshow statistic for group 1 to g-1;
        for ( i in (1:g-1)) {
        j<-(i-1)*m+1
        k<-i*m
        pred[i] <- sum(ypsort[,2][j:k])
        act[i] <- sum(ypsort[,1][j:k])
        avgp[i] <- mean(ypsort[,2][j:k])
        chi[i] <- (act[i]-pred[i])^2/(pred[i]*(1-avgp[i]))
        print(pred[i])
        print(act[i])       }
# Calculate the actual and predicted number of trees for group g;
#Calculate the average predicted survival probability for group g;
        jj<-(g-1)*m+1
        pred[g] <- sum(ypsort[,2][jj:n])
        act[g] <- sum(ypsort[,1][jj:n])
        avgp[g] <- mean(ypsort[,2][jj:n])
        chi[g] <- (act[g]-pred[g])^2/(pred[g]*(1-avgp[g]))
        print(pred[g])
        print(act[g])
#Summarize the Homser and Lemeshow statistic for group 1 to  g;
        hs<-0
        for ( i in (1:g)) { hs <-hs+chi[i] }
        print(hs)
#calculate the p-value of the Homser and Lemeshow statistic;
        pvalue <- 1-pchisq(hs,g-2)
        print(pvalue)
```

Appendix 3. SAS code for fitting the logistic model of juvenile tree mortality

This program is an example for fitting the logistic model (3.3) of juvenile tree mortality. The three data files are all from "monitor" plot data. "MOBINARY.DAT" is an individual tree data file with age, height, damage indicator, regeneration origin, and the binary response Y for each observation. "XSUMMONI.DAT" is a summary data file containing stand density and species compositions for each plot. "XIAOENVM.DAT" provides cut-block information such as regeneration treatment date and type, site class, drainage condition, and humus depth. The three data sets is merged together, so that each observation has individual tree characteristics, as well as the corresponding plot and cut-block information, listed. The first call to the procedure PROC LOGISTIC illustrates the use of stepwise selection (SELECTION=STEPWISE) to identify the variables impacting on tree survival. The second call to the procedure PROC LOGISTIC requests the maximum likelihood estimation to fit the logistic model, specifying the link function (LINK=LOGIT) and the requirement of the Hosmer-Lemeshow goodness-of-fit test (LACKFIT). The output data file "RESULT" is used for the serial correlation detection in the deviance residuals. The detection of one step serial correlation, two step serial correlation, and so on are accomplished by using the procedure PROC SQL and PROC PLOT.

```
*************************************************************************************;

* Input a data file to data TREE; each observation is a tree with the binary response y;
* PLNUMB = plot number;
* MEASNU = measurement  number;
* YEAR = measurement year;
* MONTH = measurement month;
* QPLOT = quadart number;
* TRNUMB = tree number;
* SPECIES = species code: "AW" for aspen, "SW" for white spruce, and "PL" for lodgepole pine;
* TPLANT = planting method: "B" for bareroot planting, "C" for containerized planting, "I" for ingress,
    and "A" for advanced trees;
* AGE = tree age;
* HEIGHT = tree height in cm;
* DIND = damage indicator: "H" for healthy trees, "D" for damaged trees, "M" for dead trees;
* Y = the life condition response in two years: 1 for trees still alive and 0 for trees dead;
        DATA TREE;
        INFILE "MOBINARY.DAT" LRECL=37;
```

```
        INPUT  PLNUMB 1-4 MEASNU 5 YEAR 6-7 MONTH 8-9 QPLOT 10-11 TRNUMB 12-14
        SPECIES $ 15-16 TPLANT $ 17 AGE 23-25 HEIGHT 26-28 DIND $ 35 Y 37;
*Delete advanced trees;
        IF (TPLANT='A') THEN DELETE;
*Delete already dead trees;
        IF DIND='M' THEN DELETE;
*Chose one species for analysis;
        IF SPECIES ^= 'SW' THEN DELETE;
*Introduce an indicator variable DAMAGE, and assign DAMAGE 1 for damaged trees;
        IF DIND='D' THEN DAMAGE=1; ELSE DAMAGE=0;
*Introduce two indicator variables for the planting methods: BAREROOT = 1 for bareroot planted trees,
   and CONTIANER = 1 for containerized planted trees;
        IF (TPLANT ='B') THEN BAREROOT=1; ELSE BARERROOT=0;
        IF (TPLANT ='C') THEN CONTAINER=1; ELSE CONTAINER=0;
*Change the indicator Y to Y-1, in order to fit the survival probability instead of the mortality
   probability, since the PROC LOGISTIC procedure models the probability that Y=0;
        Y=1-Y;
        RUN;
        PROC SORT;
        BY PLNUMB YEAR; RUN;
*Input a data set with stand characteristics to data SUM;
*AWDEN = number of aspen per ha;
*DENSITY = number of trees per ha;
        DATA SUM;
        INFILE "XSUMMONI.DAT" LRECL=55;
        INPUT PLNUMB 1-4 YEAR 6-7 SPECIES $ 9-10 AWDEN 12-16
            DENSITY 37-42;
        IF SPECIES^="AW" THEN DELETE;
*Calculate aspen species composition;
        SC_AW=AWDEN/DENSITY;
        RUN;
         PROC SORT;
        BY PLNUMB YEAR;  RUN;
*Merge the individual trees with the stand characteristics to data TREESUM;
        DATA TREESUM;
        MERGE TREE SUM;
        BY PLNUMB YEAR;
        IF DENSITY=. THEN DELETE;
*Calculate the inverse and logarithm of the density and aspen density;
        INVDEN=1/DENSITY;
        INVAWDEN=1/AWDEN;
        LOGDEN=LOG(DENSITY);
        LOGAW=LOG(AWDEN);
        RUN;
 *Input cut-block information to data BLOCK;
* SITECLASS = site class from phrase 3 inventory where most cut-blocks are in good and medium site;
* TRTYPE = Regeneration treatment type where 2 indicates the block was planted with containerized
   white spruce and 4 indicates it was planted with bareroot white spruce;
* DRAINAGE = an assessment of soil drainage characteristics containing 12 classes;
* HUMUS = average depth of the duff layer on the block containing 5 classes;
        DATA BLOCK;
        INFILE "XIAOENVM.DAT" LRECL=125;
        INPUT SITECLASS $ 19 PLNUMB 20-23 TRTYPE 36-37 DRAINAGE 61 HUMUS 63-64;
*Only analyze the white spruce plantations;
        IF TRTYPE^=2 AND TRTYPE^=4 THEN DELETE;
 *Assign a indicator variable SITE to present the two site classes;
```

```
        IF SITECLASS='G' THEN SITE=1; ELSE SITE=0;
*Merge the drainage classes into well drained (DW=1) and poorly drained (DW=0);
        IF (DRAINAGE=2) OR (DRAINAGE=3) OR (DRAINAGE=4) OR (DRAINAGE=9)
            THEN DW=1; ELSE DW=0;
*Merge the humus depth classes into two classes: equal to and greater than 16 cm (HS=0), and less than
  16 cm (HS=1);
        IF (HUMUS=2) OR (HUMUS=3) THEN HS=1; ELSE HS=0;
        RUN;
        PROC SORT;
        BY PLNUMB; RUN;
 *Merge individual trees and stand characteristics with cut-block information to data set ALL;
        DATA ALL;
        MERGE TREESUM BLOCK;
        BY PLNUMB;
        IF TRNUMB=. THEN DELETE;
        IF SITETYPE=. THEN DELETE;
*Calculate the logarithm of HEIGHT and AGE;
        LOGHT=LOG(HEIGHT);
        LOGAGE=LOG(AGE);
        RUN;
        PROC SORT;
        BY Y; RUN;
 *Summarize variables: HEIGHT AGE DENSITY SC_AW, by the life response in two year Y;
        PROC MEANS;
        VAR HEIGHT AGE DENSITY SC_AW;
        BY Y; RUN;
/* An example of the stepwise selection method;                                    */
        /*PROC LOGISTIC DATA=ALL;                                                  */
        /*MODEL Y = AGE LOGHT HEIGHT LOGDEN LOGAW SPCOMP BAREROOT */
        /*CONTAINER SITEG                                                          */
        /*          / LINK=LOGIT SELECTION=STEPWISE LACKFIT;                       */
        /*OUTPUT OUT=RESULT P=PS RESCHI=CHIRESID RESDEV=DEVRESID;                  */
        /*RUN;                                                                     */
*Request PROC LOGISTIC procedure on data ALL using the logistic link and following by the Homser-
  Lemeshow goodness of fit test;
        PROC LOGISTIC DATA=ALL;
        MODEL Y = AGE LOGHT INVAWDEN BAREROOT CONTAINER SITE HS DAMAGE
                / LINK=LOGIT LACKFIT;
*Output fitting results to data RESULT with predicted survival probability PS, chi-square residual
  CHIRESID, and deviance residual DEVRESID;
        OUTPUT OUT=RESULT P=PS RESCHI=CHIRESID RESDEV=DEVRESID;
        RUN;
*Calculate the deviance and the chi-squares;
        PROC SQL;
        CREATE TABLE DEVIANCE AS
        SELECT SUM(CHIRESID*CHIRESID) AS CHISQ,
            SUM(DEVRESID*DEVRESID) AS DEVIANCE,
            N(DEVRESID) AS N
        FROM RESULT;
*Request PROC SQL procedure to detect one-step serial correlation in  the deviance residuals;
        PROC SQL;
        CREATE TABLE SERIAL1 AS
        SELECT I.DEVRESID AS DEVRESID1, J.DEVRESID AS DEVRESID2
        FROM RESULT I, RESULT J
        WHERE (I.PLNUMB=J.PLNUMB) AND (I.QPLOT=J.QPLOT)
            AND (I.MEASNU=J.MEASNU-1) AND (I.TRNUMB=J.TRNUMB);
```

```
        PROC PLOT DATA=SERIAL1;
        PLOT DEVRESID2*DEVRESID1;
        RUN;
*Request PROC SQL procedure to detect two-step serial correlation in  the deviance residuals;
        PROC SQL;
        CREATE TABLE SERIAL2 AS
        SELECT I.DEVRESID AS DEVRESID1, J.DEVRESID AS DEVRESID2
        FROM RESULT I, RESULT J
        WHERE (I.PLNUMB=J.PLNUMB) AND (I.QPLOT=J.QPLOT)
            AND (I.MEASNU=J.MEASNU-2) AND (I.TRNUMB=J.TRNUMB);
        PROC PLOT DATA=SERIAL2;
        PLOT DEVRESID2*DEVRESID1;
        RUN;
*Request PROC SQL procedure to detect three-step serial correlation in  the deviance residuals;
        PROC SQL;
        CREATE TABLE SERIAL3 AS
        SELECT I.DEVRESID AS DEVRESID1, J.DEVRESID AS DEVRESID2
        FROM RESULT I, RESULT J
        WHERE (I.PLNUMB=J.PLNUMB) AND (I.QPLOT=J.QPLOT)
            AND (I.MEASNU=J.MEASNU-3) AND (I.TRNUMB=J.TRNUMB);
        PROC PLOT DATA=SERIAL1;
        PLOT DEVRESID2*DEVRESID3;
        RUN;
```

Appendix 4. SAS code for fitting the Weibull lifetime distribution model for juvenile trees

This program illustrates the nonparmatric estimation and the parametric regression model fitting for the lifetime distribution. The three data files are all from "monitor" plot data. "MOBINARY.DAT" is an individual tree data file with age, height, damage indicator, regeneration origin, censor indicator, and the observed lifetime for each observation. "XSUMMONI.DAT" is a summary data file which provides stand density and species compositions for each plot. "XIAOENVM.DAT" lists cut-block information such as regeneration treatment data and type, site class, drainage condition, and humus depth. The three data sets is merged together, so that each observation has individual tree characteristics, as well as the corresponding plot and cut-block information, listed. The call to the procedure PROC LIFETEST illustrates the nonparametric estimation for the survival function and hazard function. The product-limit method is specified (METHOD=PL), and the survival function, the log-log survival function, hazard function, and probability distribution function are required for graphic plotting (PLOTS=(S,LLS,H,P)). The call to the procedure PROC LIFEREG requests the maximum likelihood estimation for the parametric regression model, where the Weibull distribution is selected (D=WEIBULL).

```
*************************************************************************************;

* Input the individual tree data file to data TREE; each observation is a tree with observed lifetime;
* PLNUMB = plot number;
* MEASNU = measurement  number;
* YEAR = measurement year;
* MONTH = measurement month;
* QPLOT = quadart number;
* TRNUMB = tree number;
* SPECIES = species code: "AW" for aspen, "SW" for white spruce, and "PL" for lodgepole pine;
* TPLANT = planting method: "B" for bareroot planting, "C" for containerized planting, "I" for ingress,
    and "A" for advanced trees;
* AGE = tree age;
* HEIGHT = tree height in cm;
* DCODE = damage code containing 21 classes (Gilday 1990);
* DIND = damage indicator: "H" for healthy trees, "D" for damaged trees, "M" for dead trees;
```

```
* CENSOR = An indicator: 0 for trees dead before the last measurement, 1 for trees still  alive at the last
    measurement;
* LIFETIME = The elapse time from the first measurement to death or the last measurement;
        DATA TREE;
        INFILE "XMSURF.DAT" LRECL=40;
        INPUT  PLNUMB 1-4 MEASNU 5 YEAR 6-7 MONTH 8-9 QPLOT 10-11 TRNUMB 12-14
        SPECIES $ 15-16 TPLANT $ 17 AGE 23-25 HEIGHT 26-28 DCODE 32-33 DIND $ 35
        CENSOR 37 LIFETIME 39-40;
 *Delete advanced trees;
        IF (TPLANT='A') THEN DELETE;
*Delete already dead trees;
        IF DIND='M' THEN DELETE;
*Chose one species for analysis;
        IF SPECIES ^= 'SW' THEN DELETE;
*Merge damage categories into 4 classes presented by DIND: "H" for healthy trees, "S" for trees
    damaged by suppression, "C" for trees damaged by flooding and climate changes;
        IF DCODE=10 THEN DIND="C";
        IF DCODE=9 THEN DIND="C";
        IF DCODE=12 THEN DIND="S"; RUN;
        PROC SORT;
        BY PLNUMB YEAR;
*Input the data set with stand characteristics to data SUM;
*AWDEN = number of aspen per ha;
*DENSITY = number of trees per ha;
        DATA SUM;
        INFILE "XSUMMONI.DAT" LRECL=55;
        INPUT PLNUMB 1-4 YEAR 6-7 SPECIES $ 9-10 AWDEN 12-16
            DENSITY 37-42;
        IF SPECIES^="AW" THEN DELETE;
*Calculate aspen species composition;
        SC_AW=AWDEN/DENSITY; RUN;
        PROC SORT;
        BY PLNUMB YEAR;  RUN;
*Merge the individual trees with the stand characteristics to data TREESUM;
        DATA TREESUM;
        MERGE TREE SUM;
        BY PLNUMB YEAR;
        IF CENSOR=. THEN DELETE;
        IF DENSITY=. THEN DELETE;
*Calculate the inverse and logarithm of density and aspen density;
        INVDEN=1/DENSITY;
        INVAWDEN=1/AWDEN;
        LOGDEN=LOG(DENSITY);
        LOGAW=LOG(AWDEN); RUN;
*Input the cut-block information to data BLOCK;
* SITECLASS = site class from phrase 3 inventory where most cut-blocks are in good and medium site;
* TRTYPE = Regeneration treatment type where 2 indicates the block was planted with containerized
    white spruce and 4 indicates it was planted with bareroot white spruce;
* DRAINAGE = an assessment of soil drainage characteristics containing 12 classes;
* HUMUS = average depth of the duff layer on the block containing 5 classes;
        DATA BLOCK;
        INFILE "XIAOENVM.DAT" LRECL=125;
        INPUT SITECLASS $ 19 PLNUMB 20-23 TRTYPE 36-37 DRAINAGE 61 HUMUS 63-64;
*Only analyze the white spruce plantations;
        IF TRTYPE^=2 AND TRTYPE^=4 THEN DELETE;
*Assign a indicator variable SITE to present the two site classes;
```

```
        IF SITECLASS='G' THEN SITE=1; ELSE SITE=0;
*Merge the drainage classes into well drained (DW=1) and poorly drained (DW=0);
        IF (DRAINAGE=2) OR (DRAINAGE=3) OR (DRAINAGE=4) OR (DRAINAGE=9)
            THEN DW=1; ELSE DW=0;
*Merge the humus depth classes into two classes: equal to and greater than 16 cm (HS=0), and less than
  16 cm (HS=1);
        IF (HUMUS=2) OR (HUMUS=3) THEN HS=1; ELSE HS=0; RUN;
        PROC SORT;
        BY PLNUMB; RUN;
*Merge individual trees and stand characteristics with cut-block information to data set ALL;
        DATA ALL;
        MERGE TREESUM BLOCK;
        BY PLNUMB;
        IF TRNUMB=. THEN DELETE;
        IF SITETYPE=. THEN DELETE;
*Calculate the logarithm of HEIGHT and AGE;
        LOGHT=LOG(HEIGHT);
        LOGAGE=LOG(AGE); RUN;
        PROC SORT;
        BY Y; RUN;
*Summarize variables: HEIGHT AGE DENSITY SC_AW LIFETIME by CENSOR;
        PROC MEANS;
        VAR HEIGHT AGE DENSITY SC_AW LIFETIME;
        BY CENSOR; RUN;
*Request the PROC LIFETEST procedure on data ALL using the product-limit method to compute the
  nonparametric estimates of the survival distribution function and hazard function;
        PROC LIFETEST METHOD=PL NOPRINT NOTABLE PLOTS=(S, LLS, H, P)
            OUTSURV=RESULT;
*Indicate the lifetime variable and an optional censoring variable;
        TIME LIFETIME*CENSOR(1);
*Indicate TPLANT as the variable determining the strata levels for the computing;
        STRATA TPLANT; RUN;
        PROC PRINT DATA=RESULT1; RUN;
* Request the PROC LIFETREG procedure on data ALL
        PROC LIFEREG;
*Specify the class variables;
        CLASS TPLANT DIND DW HS SITE;
*Specify the model and the Weibull distribution assumption;
        MODEL LIFETIME*CENSOR(1)= HEIGHT INVAWDEN INVDEN TPLANT DIND SITE
                / D=WEIBULL;
        OUTPUT OUT=RESULT2 QUANTILES=0.5 P=PREDTIME XBETA=XBETA;
        RUN;
```

Appendix 5. SAS code for summarizing stand characteristics for SDS data

The following code gives an example for summarizing stand characteristics for nested plots, such as stand density, total basal area, average diameter, and average height. The data file "XSDS.DAT" is from Stand Dynamics data (SDS) with species, diameter at breast height, total height, and condition codes on each individual tree. Each SDS plot was a nested plot and included a tree plot, a sapling plot, and four regeneration plots (7.2.2). The variable "STATUS" in the data file indicates whether a tree belongs to the tree plot, the sapling plot, or the four regeneration plots. According to the tree "STATUS", the number of trees, total basal area, average diameter, and average height are calculated for the tree plot, the sapling plot, and the regeneration plots respectively. Since only up to 50 trees were tagged and remeasured in a regeneration plot and the remaining trees were recorded in a height class table "XSDS61.DAT", the number of trees and the average height in "XSDS61.DAT" are also summarized and added to the corresponding regeneration plots. Afterwards the stand characteristics on a ha basis are summarized with weight equal to the corresponding plot area. The summary output file is "XSDSSUM.DAT".

```
*************************************************************************************;

*Input the individual tree data file to data SDSDATA:
* PLNUMB = plot number;
* MEASNU = measurement  number;
* YEAR = measurement year;
* MONTH = measurement month;
* TRNUMB = tree number;
* SPECIES = species code: "AW" for aspen, "SW" for white spruce, and "PL" for lodgepole pine;
* DBH = diameter at breast height in cm;
* HEIGHT = total height in m;
* COND1 = condition number 1;
* COND2 = condition number 2;
* COND3 =  condition number 3
* STATUS = indicator to show which plot the tree belongs to: "T" for the tree plot, "S" for the sapling
    plot, and "R" for the regeneration plots;
        DATA SDSDATA;
        INFILE "XSDS.DAT" LRECL=64;
```

```
        INPUT GRNUMB 3-4 PLNUMB 6-7 YEAR 13-14 MONTH 15-16 MEASNU 19 SPECIES $
            24-25 TRNUMB 26-29 DBH 31-35 .2 HEIGHT 36-40 .2 COND1 41-42 COND2 43-44
            COND3 45-46 STATUS $ 64;
*Delete the dead trees;
        IF (COND1=25) OR (COND2=25) OR (COND3=25) THEN DELETE;
        IF (COND1=27) OR (COND2=27) OR (COND2=27) THEN DELETE;
        IF (COND1=29) OR (COND2=29) OR (COND3=29) THEN DELETE;
*Not that some large trees (DBH>=9.1) were not measured in  height;
        IF (DBH>0) AND (HEIGHT=0) THEN HEIGHT=.;
*Calculate basal area (cm2) for each tree;
        BA=3.1415926/4*DBH*DBH;
        RUN;
*Request PROC SQL to summarize trees in the tree plot;
        PROC SQL;
        CREATE TABLE TREES AS
        SELECT GRNUMB, PLNUMB, YEAR, AVG(MONTH) AS MONTH,
             AVG(MEASNU) AS MEASNU,
             SUM(DBH)/0.1 AS SUMDBHT, SUM(HEIGHT)/0.1 AS SUMHTT,
             N(DBH)/0.1 AS DENSITYT, SUM(BA)/0.1 AS SUMBAT
        FROM SDSDATA
        WHERE STATUS="T"
        GROUP BY GRNUMB, PLNUMB, YEAR
        ORDER BY GRNUMB, PLNUMB, YEAR;
*Request PROC SQL to summarize trees in the sapling plot;
        PROC SQL;
        CREATE TABLE SAPLING AS
        SELECT GRNUMB, PLNUMB, YEAR, AVG(MONTH) AS MONTH,
        AVG(MEASNU) AS MEASNU,
             SUM(DBH)/0.025 AS SUMDBHS, SUM(HEIGHT)/0.025 AS SUMHTS,
             N(DBH)/0.025 AS DENSITYS, SUM(BA)/0.025 AS SUMBAS
        FROM SDSDATA
        WHERE STATUS="S"
        GROUP BY GRNUMB, PLNUMB, YEAR
        ORDER BY GRNUMB, PLNUMB, YEAR;
 *Request PROC SQL to summarize trees in the four regeneration plots;
        PROC SQL;
        CREATE TABLE REGEN6 AS
        SELECT GRNUMB, PLNUMB, YEAR, AVG(MONTH) AS MONTH,
             AVG(MEASNU) AS MEASNU,
             SUM(HEIGHT)/0.004 AS SUMHTR,
             N(DBH)/0.004 AS DENSITYR
        FROM SDSDATA
        WHERE STATUS="R"
        GROUP BY GRNUMB, PLNUMB, YEAR
        ORDER BY GRNUMB, PLNUMB, YEAR;
*************************************************************************;
*Input the remaining trees ( a height class table) in the regeneration plots to data REGEN61;
* H1 = number of trees which were less than 10 cm in height;
* H2 = number of trees which were less than 20 cm and not less than 10 cm in height;
*       ……;
* H11 = number of trees which were less than 100 cm and not less than 110 cm in height;
        DATA REGEN61;
        INFILE "XSDS61.DAT" LRECL=69;
        INPUT GRNUMB 3-4 PLNUMB 6-7 YEAR 13-14 MONTH 15-16 MEASNU 19 SPECIES $
            24-25 H1 26-28 H2 30-32 H3 34-36 H4 38-40 H5 42-44 H6 46-48 H7 50-52 H8 54-56 H9
            58-60 H10 62-64 H11 66-68;
```

```
*Calculate the total number of trees;
        NUMB=H1+H2+H3+H4+H5+H6+H7+H8+H9+H10+H11;
        IF NUMB=0 THEN DELETE;
*Sum the height;
        IF NUMB ^= 0 THEN DO;
        SUMHT=0.01*(5*H1+15*H2+25*H3+35*H4+45*H5+55*H6+65*H7+75*H8+85*H9+95*H10
              +105*H11);
        END;
        RUN;
*Request PROC SQL procedure to summarize the remaining trees;
        PROC SQL;
        CREATE TABLE SUM61 AS
        SELECT GRNUMB, PLNUMB,YEAR, SUM(NUMB)/0.004 AS NUMB61,
            SUM(SUMHT)/0.004 AS SUMHT61
        FROM TEMP61
        GROUP BY GRNUMB, PLNUMB, YEAR
        ORDER BY GRNUMB, PLNUMB, YEAR;
*****************************************************************;
*Merge the summarized data from the regeneration plots and the tree height counting table to data
   REGEN;
        DATA REGEN;
        MERGE SUM61 REGEN6;
        BY GRNUMB PLNUMB YEAR;
        IF NUMB61=. THEN NUMB61=0;
        IF SUMHT61=. THEN SUMHT61=0;
        SUMHTR=SUMHTR+SUMHT61;
        DENSITYR=DENSITYR+NUMB61;
        DROP NUMB61 SUMHT61;
        RUN;
*Merge all trees together, and calculate the summary ;
        DATA SUMDEN;
        MERGE TREES SAPLING REGEN;
        BY GRNUMB PLNUMB YEAR;
        IF DENSITYT=. THEN DENSITYT=0;
        IF DENSITYR=. THEN DENSITYR=0;
        IF DENSITYS=. THEN DENSITYS=0;
        DENSITY=DENSITYT+DENSITYS+DENSITYR;
        IF SUMBAS=. THEN SUMBAS=0;
        IF SUMBAT=. THEN SUMBAT=0;
        SUMBA=SUMBAS+SUMBAT;
*CHANGE SQUARE CM TO SQUARE METER;
        SUMBA=0.0001*SUMBA;
        SUMBAT=0.0001*SUMBAT;
        SUMBAS=0.0001*SUMBAS;
        IF SUMDBHS=. THEN SUMDBHS=0;
        IF SUMDBHT=. THEN SUMDBHT=0;
        IF (DENSITYS=0) AND (DENSITYT=0) THEN AVGDBH=.;
        IF (DENSITYS ^= 0) OR (DENSITYT ^= 0) THEN
              AVGDBH=(SUMDBHS+SUMDBHT)/(DENSITYS+DENSITYT);
        IF SUMHTS=. THEN SUMHTS=0;
        IF SUMHTT=. THEN SUMHTT=0;
        IF SUMHTR=. THEN SUMHTR=0;
        IF DENSITY=0 THEN AVGHT=.;
        IF DENSITY ^= 0 THEN AVGHT=(SUMHTS+SUMHTT+SUMHTR)/DENSITY;
        IF DENSITYR=0 THEN AVGHTR=.;  ELSE AVGHTR=SUMHTR/DENSITYR;
        IF DENSITYS=0 THEN AVGDBHS=.; ELSE AVGDBHS=SUMDBHS/DENSITYS;
```

```
        IF DENSITYS=0 THEN AVGHTS=.; ELSE AVGHTS=SUMHTS/DENSITYS;
        IF DENSITYT=0 THEN AVGDBHT=.; ELSE AVGDBHT=SUMDBHT/DENSITYT;
        IF DENSITYT=0 THEN AVGHTT=.; ELSE AVGHTT=SUMHTT/DENSITYT;
        RUN;
*Output the summary data set to "XSDSSUM.DAT";
        DATA FINAL;
        SET SUMDEN;
        FILE "XSDSSUM.DAT" LRECL=101;
        PUT GRNUMB 1-2 PLNUMB 3-4 YEAR 6-7 MONTH 8-9 MEASNU 11 DENSITYR 13-19 .1
            AVGHTR 20-24 .2 DENSITYS 25-31 .1 AVGDBHS 32-35 .1 AVGHTS 36-40 .2 SUMBAS
            41-49 .4 DENSITYT 50-56 .1 AVGDBHT 58-61 .1 AVGHTT 62-66 .2 SUMBAT 67-75 .4
            DENSITY 76-83 .1 AVGDBH 84-87 .1 AVGHT 88-92 .2 SUMBA 93-101 .4;
        RUN;
```

Appendix 6. SAS code for fitting the diameter increment model

The following code illustrates the fitting procedure of the diameter increment model for white spruce. The three data files "DIPSPSW.DAT", "XSDSINSW.DAT", and "XPGINCSW.DAT" were summarized from original PSP, SDS, and RPSP data. Each observation is an individual white spruce tree with the measurements of diameter, diameter increment, height, total basal area of larger trees, and the corresponding stand characteristics such as stand basal area, density, species composition, site productivity, average diameter, and average height. The three data files are put together to provide a sufficient data for fitting the diameter increment model. The call to the procedure PROC REG requests the maximum R^2 improvement technique for the explanatory variables selection in the linearized diameter increment model (7.3). The call to the procedure PROC NLIN illustrates the weighted nonlinear regression for model (7.4). The initial estimates of parameters are given (PARMS A1, …, C2), and the partial derivatives of the model are listed (DER.A1, …DER.C2). The weight is provided (_Weight_). The output data file "RESULT" from the nonlinear regression is used for residual analysis and serial correlation detection. The detection of one step serial correlation, two step serial correlation, and so on are accomplished by using the procedure PROC SQL and PROC PLOT.

```
*************************************************************************************;

* Input spruce trees with stand characteristics from PSP to data PSPSW;
* PLNUMB = plot number;
* MEASNU = measurement  number;
* YEAR = measurement year;
* MONTH = measurement month;
* TRNUMB = tree number;
* SPECIES = species code: "AW" for aspen, "SW" for white spruce, and "PL" for lodgepole pine;
* DBH = diameter at breast height in cm;
* HEIGHT = total height in m;
* GGR = total basal area per  ha of trees greater than the subject tree (m2);
* DIN = annual diameter increment (cm);
* HTINC = annual height increment (m);
* AVGDBH = average diameter;
* AVGHT = average height;
* BA = total basal area per ha (m2);
* DENSITY = number of trees per ha;
```

```
* SWDEN = number of white spruce per ha;
* SC_SW = white spruce composition based on basal area;
* AWDEN = number of aspen per ha;
* SC_AW = aspen composition based on basal area;
* SPI = site productivity index;
        DATA PSPSW;
        INFILE "DIPSPSW.DAT" LRECL=175;
        INPUT GRNUMB 3-12 PLNUMB 13 MEASNU 14-15 YEAR 16-17 MONTH 18-20 .1
            TRNUMB 21-24 SPCODE $ 25-26 DBH 27-32 .1 HEIGHT 33-38 .1 GGR 46-53 .4 DIN 59-
            63 .2 HTINC 64-68 .2 AVGDBH 69-73 .1 AVGHT 74-78 .1 BA 79-86 .4 DENSITY 87-92
            SWDEN 93-97 SC_SW 98-102 .2 AWDEN 103-107 SC_AW 108-112 .2 SPI 150-154 .1;
        RUN;
* Input spruce trees with stand characteristics from SDS to data SDSSW;
* GRNUMB = group number;
* PLNUMB = plot number;
* MEASNU = measurement  number;
* YEAR = measurement year;
* MONTH = measurement month;
* TRNUMB = tree number;
* SPECIES = species code: "AW" for aspen, "SW" for white spruce, and "PL" for lodgepole pine;
* DBH = diameter at breast height in cm;
* HEIGHT = total height in m;
* GGR = total basal area per  ha of trees greater than the subject tree (m2);
* DIN = annual diameter increment (cm);
* HTINC = annual height increment (m);
* AVGDBH = average diameter;
* AVGHT = average height;
* BA = total basal area per ha (m2);
* DENSITY = number of trees per ha;
* SWDEN = number of white spruce per ha;
* SC_SW = white spruce composition based on basal area;
* AWDEN = number of aspen per ha;
* SC_AW = aspen composition based on basal area;
* SITE = site class in the phrase 3 inventory;
        DATA SDSSW;
        INFILE "XSDSINSW.DAT" LRECL=140;
        INPUT RETYPE 1 GRNUMB 3-4 PLNUMB 6-7 YEAR 13-14 MONTH 15-17 .1 MEASNU 19
            SPECIES $ 24-25 TRNUMB 26-29 DBH 31-35 .2 HEIGHT 36-40 .2 GGR 53-60 .4 DIN 65-
            70 .2 HTINC 71-76 .2 DENSITY 77-84 .1 AVGDBH 85-88 .1 AVGHT 89-93 .2 BA 94-102
            .4 SWDEN 103-110 .1 SC_SW 111-115 .2 AWDEN 121-128 .1 SC_AW 129-133 .2 SITE $ 139;
*Enlarge the group number in order to differentiate it from the group numbers in PSP.
        GRNUMB=1000+GRNUMB;
*Convert the site class in the phrase 3 inventory to the site productivity index;
        IF SITE='G' THEN SPI=22;
        IF SITE='M' THEN SPI=18;
        IF SITE='F' THEN SPI=14; *rare case;
        RUN;
* Input white spruce trees with stand characteristics from RPSP to data RPSPSW;
* LOCATION = plot location in column 1-8;
* MEASNU = measurement  number;
* YEAR = measurement year;
* MONTH = measurement month;
* TRNUMB = tree number;
* SPECIES = species code: "AW" for aspen, "SW" for white spruce, and "PL" for lodgepole pine;
* DBH = diameter at breast height in cm;
* HEIGHT = total height in m;
```

```
* GGR = total basal area per  ha of trees greater than the subject tree (m2);
* DIN = annual diameter increment (cm);
* HTINC = annual height increment (m);
* AVGDBH = average diameter;
* AVGHT = average height;
* BA = total basal area per ha (m2);
* DENSITY = number of trees per ha;
* SWDEN = number of white spruce per ha;
* SCSW = white spruce composition based on basal area;
* AWDEN = number of aspen per ha;
* SCAW = aspen composition based on basal area;
* SITE = site class in the phrase 3 inventory;
        DATA RPSPSW;
        INFILE "XPGINCSW.DAT" LRECL=150;
        INPUT LO 1 CA 3-5 TION 7-8 PLNUMB 10-11 YEAR 15-16 MONTH 17-19 .1 MEASNU 25
            TRNUMB 27-29 SPECIES $ 32-33 DBH 35-38 .1 HEIGHT 40-44 .2 GGR 63-69 .4 DIN 74-
            79 .2 HTINC 80-85 .2 DENSITY 87-94 .1 AVGDBH 95-98 .1 AVGHT 99-103 .2 BA 104-
            112 .4 SWDEN 113-120 .1 SCSW 121-125 .2 AWDEN 131-138 .1 SCAW 139-143 .2 SITE $
            149;
*Convert plot location to a group number which is related to location and different from the group
   numbers ofr PSPSW and SDSSW data;
        GRNUMB=TION+CA*100+LO*100000;
*Convert the site class in the phrase 3 inventory to the site productivity index;
        IF SITE='G' THEN SPI=22;
        IF SITE='M' THEN SPI=18;
        IF SITE='F' THEN SPI=14;   *rare case;
        RUN;
*Include the three data sets together to SWDIN;
        DATA SWDIN;
        SET PSPSW SDSSW RPSPSW;
*Delete dead trees;
        IF DBH=0 THEN DELETE;
        IF DBH=. THEN DELETE;
*Only include trees with height measurement;
        IF HEIGHT=. THEN DELETE;
*Compute possible variable transformations;
        DSQUARE=DBH*DBH;
        LOGDBH=LOG(DBH);
        LOGHT=LOG(HEIGHT);
        LOGDIN=LOG(DIN);
        LOGDEN=LOG(DENSITY);
        LOGBA=LOG(BA);
        LOGSPI=LOG(SPI);
        RUN;
/* An example for fitting the linearized model with the maximum R2 improvement technique for variable
   selection (7.3);                                                                          */
/* PROC REG DATA=SWDIN;                                                                      */
/* MODEL LOGDIN = LOGDBH DSQUARE DBH HEIGHT BA GGR DENSITY                                   */
/*                        SWDEN AWDEN SCSW SCAW  CDSPI LOGSPI                                */
/*                / SELECTION=MAXR;                                                          */
*Request PROC NLIN procedure on data SWDIN using Marquardt method;
        PROC NLIN DATA=SWDIN METHOD=MARQUARDT;
*Set the initial values for parameters;
        PARMS A1=0.343214 A2=-0.1790372 A3=-0.02012508 A4=0.01814169
              A5=-0.0845195 A6=-0.49858168 B1=0.34706048
              C1=-0.00014689 C2=0.0000;
```

```
*Calculate parameters as defined in (7.10);
        A=A1*(BA)**A2*EXP(A3*BA+A4*SPI+A5*GGR+A6*SC_SW);
        B=B1;
        C=C1+C2/HEIGHT;
*Specify the diameter increment base curve in (7.10);
        MODEL DIN = A*(DBH)**B*EXP(C*DBH*DBH);
*Calculate the partial derivative of the model (7.10) to each parameter;
        DER.A1=(BA)**A2*EXP(A3*BA+A4*SPI+A5*GGR+A6*SC_SW)*(DBH)**B*EXP(C*DBH*D
            BH);
        DER.A2=LOG(BA)*A*(DBH)**B*EXP(C*DBH*DBH);
        DER.A3=BA*A*(DBH)**B*EXP(C*DBH*DBH);
        DER.A4=SPI*A*(DBH)**B*EXP(C*DBH*DBH);
        DER.A5=GGR*A*(DBH)**B*EXP(C*DBH*DBH);
        DER.A6=SC_SW*A*(DBH)**B*EXP(C*DBH*DBH);
        DER.B1=LOG(DBH)*A*(DBH)**B*EXP(C*DBH*DBH);
        DER.C1=DBH*DBH*A*(DBH)**B*EXP(C*DBH*DBH);
        DER.C2=DBH*DBH/HEIGHT*A*(DBH)**B*EXP(C*DBH*DBH);
*Assign the weight;
        _WEIGHT_=DBH**0.2;
*Output the predicted diameter increments PDIN, residuals RESID, and studentized residuals STRESID ;
        OUTPUT OUT=RESULT P=PDIN R=RESID STUDENT=STRESID;
        RUN;
*Plot scatter plots for residual examinations;
        PROC PLOT DATA=RESULT;
        PLOT DIN*DBH='A' YHAT*DBH='P' / OVERLAY;
        PLOT STRESID*PDIN;
        PLOT RESID*DBH;
        PLOT RESID*HEIGHT;
        PLOT RESID*BA;
        PLOT RESID*GGR;
        RUN;
*Request PROC SQL procedure to detect one-step serial correlation in  the residuals;
        PROC SQL;
        CREATE TABLE SERIAL1 AS
        SELECT I.RESID AS RESID1, J.RESID AS RESID2
        FROM RESULT I, RESULT J
        WHERE (I.GRNUMB=J.GRNUMB) AND (I.PLNUMB=J.PLNUMB)
            AND (I.MEASNU=J.MEASNU-1) AND (I.TRNUMB=J.TRNUMB);
        PROC PLOT DATA=SERIAL1;
        PLOT RESID2*RESID1;
        RUN;
*Request PROC SQL procedure to detect two-step serial correlation in  the residuals;
        PROC SQL;
        CREATE TABLE SERIAL2 AS
        SELECT I.RESID AS RESID1, J.RESID AS RESID2
        FROM RESULT I, RESULT J
        WHERE (I.GRNUMB=J.GRNUMB) AND (I.PLNUMB=J.PLNUMB)
            AND (I.MEASNU=J.MEASNU-2) AND (I.TRNUMB=J.TRNUMB);
        PROC PLOT DATA=SERIAL2;
        PLOT RESID2*RESID1;
        RUN;
*Request PROC SQL procedure to detect three-step serial correlation in  the residuals;
        PROC SQL;
        CREATE TABLE SERIAL3 AS
        SELECT I.RESID AS RESID1, J.RESID AS RESID2
        FROM RESULT I, RESULT J
```

```
WHERE (I.GRNUMB=J.GRNUMB) AND (I.PLNUMB=J.PLNUMB)
    AND (I.MEASNU=J.MEASNU-3) AND (I.TRNUMB=J.TRNUMB);
PROC PLOT DATA=SERIAL3;
PLOT RESID2*RESID1;
RUN;
```

B45378

University of Alberta Library

0 1620 0713 7894